

The Adenoviruses

Edited by

HAROLD S. GINSBERG

*College of Physicians and Surgeons of Columbia University
New York, New York*

PLENUM PRESS • NEW YORK AND LONDON

Library of Congress Cataloging in Publication Data

Main entry under title:

The Adenoviruses.

(The Viruses)

Includes bibliographical references and index.

1. Adenoviruses. I. Ginsberg, Harold S., 1917- . II. Series.

QP396.A34 1984

576'.64

84-8264

ISBN 0-306-41592-5

© 1984 Plenum Press, New York
A Division of Plenum Publishing Corporation
233 Spring Street, New York, N.Y. 10013

All rights reserved

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the Publisher

Printed in the United States of America

CHAPTER 3

The Structure of the Genome

JOHN S. SUSSENBACH

I. INTRODUCTION

Adenovirus particles have a highly ordered structure and are composed of protein and DNA. Human adenoviruses contain about 87% protein and 13% DNA (Green and Piña, 1963), while the larger avian chick embryo lethal orphan (CELO) virus consists of 83% protein and 17% DNA (Laver *et al.*, 1971). In virions, the viral DNA is tightly associated with several virus-coded proteins. Disruption of virions with acetone, urea, or pyridine, or repeated freezing and thawing, releases the viral cores, which, in addition to the viral DNA, still contain about 18–20% of the total protein of the virions (Laver *et al.*, 1967, 1968; Maizel *et al.*, 1968; Prage *et al.*, 1968, 1970). The proteins found in viral cores are mainly two basic polypeptides. The major core protein is identical to polypeptide VII [molecular weight 18,000 (18K)], of which about 1000 copies are present in each viral particle. The minor core protein is polypeptide V (molecular weight 45.5K), of which each virion contains about 200 copies (Laver *et al.*, 1968; Prage *et al.*, 1968, 1970; Prage and Pettersson, 1971; Russell *et al.*, 1971; Everitt *et al.*, 1973; Laver, 1970). However, when cores are prepared by extraction of virions with sarkosyl, only polypeptide VII is found associated with the DNA (Brown *et al.*, 1975). The different protein compositions of pyridine and sarkosyl cores suggest that polypeptide VII is more intimately associated with the viral genome than is polypeptide V.

Corden *et al.* (1976) concluded that adenovirus DNA packed in virions has a chromatinlike structure. They found that digestion of disrupted virions with micrococcal nuclease cleaves the viral genome into fragments about 200 nucleotides long. However, these experiments could

JOHN S. SUSSENBACH • Laboratory for Physiological Chemistry, State University of Utrecht, 3521 GG Utrecht, The Netherlands.

not be repeated by Tate and Philipson (1979). Mirza and Weber (1982) proposed that although adenovirus DNA is indeed packed into subunits, its organization in the virion is not completely the same as that of eukaryotic chromatin. Partial deoxyribonuclease (DNase) digestion of eukaryotic chromatin leads to stretches of DNA with a length of 200 nucleotide pairs associated with histones. Mirza and Weber (1982) found that viral chromatin does indeed have a nucleosomelike structure, but that partial DNase digestion yields monomers of about 150 nucleotide pairs of DNA wrapped around three dimers of polypeptide VII. These monomers are linked by a variable length of DNA associated with one copy of polypeptide V.

Since adenovirus DNA is tightly associated with virion proteins, protein-free DNA can be obtained only by extensive digestion of virions or viral cores with proteolytic enzymes (papain, pronase, or proteinase K) followed by sodium dodecyl sulfate (SDS)-phenol extraction (van der Eb and van Kesteren, 1966; Green *et al.*, 1967; van der Eb *et al.*, 1969; Laver *et al.*, 1971). The DNA thus isolated has a linear structure and has been characterized in great detail.

An alternative isolation procedure for adenovirus DNA was first applied by Bellett and co-workers for CELO and adenovirus type 2 (Ad2) DNA (Robinson *et al.*, 1973; Robinson and Bellett, 1975a). These investigators isolated DNA in the absence of proteolytic enzymes, employing an extraction with 4 M guanidinium hydrochloride. The isolated DNA has in the electron microscope (EM) a circular structure, which can be converted into a linear configuration by digestion of the preparation with proteolytic enzymes (Robinson *et al.*, 1973). Similar studies have also been performed for Ad5 DNA (Keegstra *et al.*, 1977). The sensitivity of the circular structures for proteolytic enzymes suggests that the circular structures are maintained by a protein linker.

By *in vitro* labeling of the protein moiety with ^{125}I , it could be demonstrated that a polypeptide with a molecular weight of 55K is covalently attached to the 5' end of each DNA strand (Rekosh *et al.*, 1977). This protein, designated terminal protein, has a hydrophobic character, which facilitates joining of the ends of the DNA-protein complexes, resulting in the formation of circular structures and concatemers. The properties of the linear deproteinized DNA as well as the characteristics of the circular DNA-protein complexes are discussed in more detail in the following sections.

II. GROUPING OF ADENOVIRUSES BASED ON DNA HOMOLOGY

The different human adenoviruses have been classified into subgroups on the basis of different criteria. Rosén (1960) originally proposed three subgroups based on differences in hemagglutinating capacity.

Hierholzer (1973) extended this classification system to ten subgroups. On the basis of the apparent molecular weights of virion polypeptides V, VI, and VII, Wadell (1978) arranged 20 human serotypes into five groups. A completely different type of classification is based on the oncogenicity of the human adenoviruses. The different serotypes have been subdivided into a highly oncogenic subgroup A (Ad12, Ad18, Ad31), a weakly oncogenic subgroup B (e.g., Ad3 and Ad7), and a nononcogenic subgroup C (e.g., Ad2 and Ad5) (Trentin *et al.*, 1962; Girardi *et al.*, 1964; Huebner *et al.*, 1962, 1965; Larson *et al.*, 1965; Pereira *et al.*, 1965; Green, 1970). It is interesting to note that there is a correlation between the guanine-cytosine (GC) content of the human adenovirus DNAs and the oncogenicity of the viruses. The GC content of the DNAs decreases with increasing oncogenicity (Piña and Green, 1965) (Table I). Probably this correlation has no physiological basis, since, in contrast to the human adenoviruses, the oncogenic simian adenoviruses tend to have slightly higher GC contents than the nononcogenic adenoviruses (Goodhearst, 1971). Further, the oncogenic simian serotypes have GC contents that are in general higher than those of the nononcogenic human serotypes.

The most meaningful and fundamental way to group adenoviruses is based on DNA sequence homology. Fortunately, the DNA homology grouping is in agreement with other groupings of human adenoviruses on the basis of oncogenicity, GC content, and molecular characteristics of viral proteins (Table I). Originally, Green *et al.* (1970) determined the homology among different DNAs employing filter hybridization. Recently, the classification was improved by employment of liquid-phase molecular hybridization with *in vitro*-labeled viral DNA. A total of 31 different human adenovirus serotypes were divided into five different subgroups, A-E (Green *et al.*, 1979b). In general, members of the same subgroup have genomes that are homologous for more than 90%. However, members of subgroup A share only 48–69% of their DNA sequences. The homology among members of different subgroups is less than 20% (Table I).

The major regions of least homology among DNAs of different human serotypes have been visualized by heteroduplex mapping (Garon *et al.*, 1973). Heteroduplexes of subgroups B and C DNAs contain two major regions of heterology located at positions 50–65 and 78–91 on the adenovirus genome map. Heteroduplexes of members of subgroup A show a more complex distribution of homologous and heterologous regions. However, in this case, too, heterology is found at the two positions mentioned above.

Using the single-strand specific endonuclease from *Neurospora crassa*, Bartok *et al.* (1974) were able to digest specifically the heterologous regions from heteroduplexes of Ad2 and Ad5 DNA and obtained three specific fragments, in agreement with the heteroduplex mapping. The heterologous regions contain the genetic information of the major coat proteins hexon and fiber, which play an important role in the se-

TABLE I. Properties of Human Adenovirus DNA Homology Groups A-E^a

Groups	Types	DNA homology ^b	DNA MWs ($\times 10^{-6}$) ^c	DNA GC [%] ^d	ITR length ^e	Tumor induction ^f	Cell transformation ^g	HA group ^h
A	12, 18, 31	48-69% within group, 8-20% with other types	19.2-22.0	47-49	162/164 for Ad12, 165 for Ad18	High	+	3B
B	3, 7, 11, 14, 16, 21	89-94% within group, 9-20% with other types	22.7-23.0	49-52	136 for Ad3 and Ad7	Weak	+	1A, B
C	1, 2, 5, 6	99-100% within group, 10-16% with other types	23.0	57-59	102/103 for Ad2, 103 for Ad5	Nil	+	3A
D	8-10, 13, 15, 17, 19, 20, 22-30, 32, 33, 36, 37	94-99% within group, 4-17% with other types	?	57-59	?	Nil	?	2A-F
E	4	4-23% with other types	22.8	High	116	Nil	?	3A

^a [MW] Molecular weight, [ITR] inverted terminal repetition, [HA] hemagglutination.^b Data from Green *et al.* (1979b).^c Data from Green and Piña (1964), van der Eb and van Kesteren (1966), and Tibbetts (1977).^d Data from Piña and Green (1965).^e Data from Steenbergh *et al.* (1977), Arrand and Roberts (1979), Shinagawa and Padmanabhan (1979), Tokunaga *et al.* (1982), Sugisaka *et al.* (1980), Caron *et al.* (1982), and Schwarz *et al.* (1982).^f Data from Girardi *et al.* (1964), Huebner *et al.* (1962, 1965), Larson *et al.* (1965), Percira *et al.* (1965), and Trentine *et al.* (1962).^g Data from McBride and Weiner (1964), Sekikawa *et al.* (1978), Freeman *et al.* (1967), Gallimore (1974), van der Eb *et al.* (1977), and McAllister *et al.* (1969).^h Data from Hierholzer (1973).

rological classification of the different adenovirus serotypes. In addition, one of the heterologous regions codes for a group of nonvirion early proteins (see Section VII).

III. PHYSICOCHEMICAL PROPERTIES OF ADENOVIRUS DNA

DNA, extracted from adenovirus particles employing digestion with proteolytic enzymes, has a linear double-stranded structure (van der Eb and van Kesteren, 1966; Green *et al.*, 1967; van der Eb *et al.*, 1969; Younghusband and Bellett, 1971). The size of the viral genome varies from serotype to serotype. The molecular weights of the human adenovirus DNAs range from $19\text{--}22 \times 10^6$ for the highly oncogenic serotypes Ad12, Ad18, and Ad31 to $23\text{--}24 \times 10^6$ for the nononcogenic serotypes Ad1, Ad2, and Ad5 (Green *et al.*, 1967) (Table I). On the basis of nucleotide sequence data and the sum of restriction fragments, it has been inferred that the genome of Ad2 and Ad5 is about 36,000 nucleotide pairs and that Ad12 DNA is 34,300 nucleotide pairs long. The sizes of the genomes of nonhuman serotypes are comparable to those of their human counterparts [that of mouse serotype FL DNA being 20.7×10^6 (Temple *et al.*, 1981) and of simian adenovirus SA7 DNA being 22×10^6 (Burnett and Harrington, 1968)]. On the other hand, the genome of the avian chick embryo lethal orphan (CELO) virus is much larger, measuring 30×10^6 (Younghusband and Bellett, 1971; Laver *et al.*, 1971).

When native adenovirus DNA is digested with *Escherichia coli* exonuclease III and is subsequently examined under the EM, no circularization of the linear genome is observed, indicating that adenovirus DNA is not terminally redundant as T7 DNA (Green *et al.*, 1967; Younghusband and Bellett, 1971). On the other hand, when double-stranded DNA (dsDNA) is denatured and reannealed at low DNA concentrations, both strands of human as well as of avian adenovirus DNA are able to form single-stranded circles (Garon *et al.*, 1972; Wolfson and Dressler, 1972; Robinson and Bellett, 1975b). The formation of single-stranded circles indicates that adenovirus DNA contains an inverted terminal repetition. This inverted terminal repetition is discussed in more detail in Section V.

The distribution of adenine–thymine (AT) and GC base pairs in adenovirus DNA has been investigated by partial thermal denaturation mapping. The unique thermal denaturation patterns of DNAs from Ad2, Ad5, and Ad12, the avian CELO virus, and the mouse strain FL indicate that adenovirus DNA is not circularly permuted as T7 DNA, but that all DNA molecules from the same serotype have an identical nucleotide sequence (Doerfler and Kleinschmidt, 1970; Younghusband and Bellett, 1971; Doerfler *et al.*, 1972; Ellens *et al.*, 1974; Temple *et al.*, 1981). In most denaturation patterns, the distribution of AT and GC base pairs

along the DNA molecule is asymmetrical. By convention, the AT-rich half of an adenovirus DNA molecule has been designated the right-hand half of the molecule (Doerfler and Kleinschmidt, 1970). In some cases (Ad2 and Ad5), the AT- and GC-rich halves of the DNA molecules can be separated by CsCl or HgCl₂-Cs₂ SO₄ gradient centrifugation of sheared DNA (Kimes and Green, 1970; Doerfler and Kleinschmidt, 1970; Horwitz, 1974; Graham *et al.*, 1974b). However, due to the more even distribution of AT and GC base pairs in Ad12 DNA, separation of the left and right halves of Ad12 DNA by this procedure is not possible (Doerfler *et al.*, 1972).

Separation of the complementary strands of adenovirus DNA can be performed by complexing of the single strands of denatured native DNA with poly(I:G) or poly(U:G). Intact complementary strands have been obtained for Ad2, Ad5, Ad7, and Ad12 DNA (Kubinski and Rose, 1967; Landgraf-Leurs and Green, 1971; Patch *et al.*, 1972; Tibbetts *et al.*, 1974; Vlak *et al.*, 1975). Since the two complementary strands bind unequal amounts of the copolymers, the two strands can be separated by equilibrium density-gradient centrifugation or by gel electrophoresis (Goldbach *et al.*, 1978). Complementary strands of Ad2 and Ad5 DNA have also been separated by alkaline CsCl equilibrium density-gradient centrifugation (Sussenbach *et al.*, 1973; Sharp *et al.*, 1975). The buoyant densities of the two strands in alkaline CsCl differ by 2–4 mg/ml, which is sufficient for separation. The heavy strands of Ad2 and Ad5 DNA obtained by poly(U:G)-CsCl gradient centrifugation have the lower density in alkaline CsCl (Tibbetts *et al.*, 1974; Vlak *et al.*, 1975).

Tibbetts *et al.*, (1973) showed that Ad2 single-stranded DNA (ssDNA) is retained by hydroxyapatite columns under conditions generally used for selective retention of dsDNA, probably due to partially complementary regions in the single strands. Other indications for regions of complementarity in adenovirus ssDNA were obtained by EM. Under suitable conditions, an extended region of secondary structure is observed at position 73 on the conventional adenovirus map (Wu *et al.*, 1977). Regions that contain complementary sequences were also detected at the molecular termini (Padmanabhan and Green, 1976; Wu *et al.*, 1977). Digestion of native Ad2 DNA with exonuclease III followed by repair synthesis of the exposed single-stranded ends with DNA polymerase I revealed the presence of self-complementary sequences about 50 nucleotides long, located at a distance of about 180 nucleotides from each molecular end (Padmanabhan and Green, 1976). Nucleotide sequence analysis of the termini confirmed the existence of self-complementary sequences in these regions.

IV. COORDINATE SYSTEM

To come to an unambiguous nomenclature for the two complementary strands of adenovirus DNA, it has been proposed to adopt a nomen-

clature that is based on the direction of transcription, rather than on physical properties, e.g., densities. By convention, the AT-rich half of the DNA molecule is oriented to the right and the strand transcribed to the right is called the r-strand, while the leftward-transcribed strand is designated the l-strand.* The r-strand appears to be identical to the strand with the higher density in alkaline CsCl and to the strand with lower density in poly(U:G)-CsCl (see the proposal in *J. Virol.* 22:830, 1977). Further, it is agreed to divide the adenovirus DNA into 100 map units (m.u.) from left to right on the viral genome.

The agreement on a unique orientation of adenovirus DNA molecules formed the basis for an unambiguous mapping of significant landmarks on the adenovirus genome. With the discovery and the purification of restriction endonucleases, powerful tools became available to dissect the adenovirus genome in distinct specific fragments (for a review of available enzymes, see Roberts, 1981). These fragments have been used to unravel the organization of the adenovirus genome in detail. For many adenovirus serotypes, accurate restriction endonuclease cleavage maps of the viral genome are available, and with the increasing knowledge of the nucleotide sequences of several adenovirus DNAs, this number is still growing. A summary of restriction endonuclease cleavage maps is presented in Appendix A.

Many restriction fragments have been inserted into prokaryotic plasmids employing recombinant DNA techniques (Stenlund *et al.*, 1980). These adenovirus DNA-containing plasmids are very useful for obtaining large amounts of specific fragments, especially of poorly growing serotypes. They have frequently been used for nucleotide sequence analysis and site-directed mutagenesis. The two complementary strands of restriction fragments have been separated by annealing denatured fragments in the presence of an excess of one of the intact complementary strands followed by separation of the partial duplex and the remaining single strand. Strand separation has also been obtained by gel electrophoresis of denatured restriction fragments (Tibbetts and Pettersson, 1974; Sharp *et al.*, 1975; Sussenbach *et al.*, 1973; Goldbach *et al.*, 1978). These single strands have frequently been used to isolate specific messenger RNA (mRNA) species.

The most detailed information on the structure of the adenovirus genome and the positions of important landmarks became available by nucleotide sequence analysis of DNAs from different adenovirus serotypes (see Appendix B). The most extended sequences have been established for Ad2 DNA, of which about 70% has been sequenced (Arrand and Roberts, 1979; Zain and Roberts, 1979; Zain *et al.*, 1979a,b; Shinagawa and Padmanabhan, 1979; Galibert *et al.*, 1979; Akusjärvi and Pettersson, 1978a,b, 1979a,b; Hérisse *et al.*, 1980, 1981; Akusjärvi *et al.*,

It should be noted that r-strand transcripts are equivalent to l-strand DNA sequences and that l-strand transcripts are homologous to r-strand sequences.

1980, 1981; Shinagawa *et al.*, 1980; Hérisse and Galibert, 1981; Aleström *et al.*, 1980, 1982; Akusjärvi and Persson, 1981a; Kruijer *et al.*, 1982; Gingeras *et al.*, 1982). This allows the positioning of many landmarks on the Ad2 genome at the nucleotide level. Comparison of the Ad2 nucleotide sequence and the restriction maps revealed that the nucleotide equivalent of 1% of the genome depends on the particular location on the Ad2 genome (Gingeras *et al.*, 1982). It was derived that a value of 365 nucleotides for 1% gives the best fit for the left end, while a value of 357 nucleotides for 1% is the best fit for the right end. The differences in nucleotide equivalent for 1% are probably caused by the differences in nucleotide composition between the right and left halves of the Ad2 genome.

V. INVERTED TERMINAL REPETITION

The existence of an inverted terminal repetition (ITR) in adenovirus DNA was discovered when denatured DNA was reannealed at low concentrations and examined under the EM. A high percentage of the single strands were present in a circular form, indicating that adenoviral DNA contains an ITR (Garon *et al.*, 1972; Wolfson and Dressler, 1972). So far, ITRs have been detected in every serotype investigated, although the length of the repetitions may vary (Table I). The general occurrence of an ITR in adenovirus DNA suggests very strongly that this feature plays an important role in viral propagation.

The single-stranded circular structures have a rather high thermal stability, which is consistent with a highly ordered base-pairing between the terminal sequences (Garon *et al.*, 1972; Wolfson and Dressler, 1972). It also suggests that the ITRs must be of considerable length. Circularization of adenovirus ssDNA can be abolished by digestion with exonuclease III, and this treatment has been used to estimate the size of the terminal repetitions. Garon *et al.* (1972) concluded that the length of the terminal repetition ranged from 350 base pairs (bp) for Ad2 to 1400 bp for Ad31. However, since inverted repeats of these sizes can be visualized under the EM and no double-stranded regions were detected in the single-stranded circles, it was concluded that the exonuclease III experiments obviously lead to an overestimation of the lengths of the ITRs. An exceptionally long ITR was detected in Ad18 DNA (Garon *et al.*, 1975). In single-stranded circles of this serotype, a double-stranded panhandle with a mean length of 0.31 μm was seen, equivalent to 3% of the genome length.

A more accurate estimate of the size of the ITR of Ad2 DNA was obtained by restriction enzyme analysis of end-labeled DNA. When a restriction enzyme cleaves within the repeated sequence, both molecular ends will yield a fragment of the same size, while cleavage outside the repeated sequence will yield fragments of different size. Employing this

approach, Roberts *et al.* (1974) estimated that the terminal repetition of Ad2 DNA is between 100 and 140 nucleotides long (also see Arrand *et al.*, 1975).

Recently, nucleotide sequence analysis has been used to determine exactly the size and composition of several adenovirus serotypes (Appendix B). Some general features of the adenovirus ITRs can be demonstrated in the ITR of Ad5 DNA, the first sequenced repetition. The ITR of Ad5 is 103 bp long (Steenbergh *et al.*, 1977). Its sequence is unique and does not contain extended self-complementary regions. A striking property of the Ad5 terminal repetition is the asymmetrical distribution of GC and AT base pairs. The first 50 bp contain 72% AT, while the next 50 bp have only 27% AT. Although the lengths of inverted repeats of other serotypes may differ considerably, they all show the same asymmetrical distribution of base pairs. As for a function of this property, it is not unlikely that the high AT content of the first half of terminal repetitions is of relevance for a rapid unwinding of the molecular ends during initiation of DNA replication.

Comparison of the inverted repetitions of serotypes from the same subgroup shows a high degree of homology (see Appendix B). The repetitions of Ad2 and Ad5 both have a length of 103 bp and are completely identical (Steenbergh *et al.*, 1977; Shinagawa and Padmanabhan, 1979), although the repetition of a particular Ad2 strain has been described that is 102 bp long (Arrand and Roberts, 1979). The terminal repetitions of Ad3 and Ad7 strain Greider both have a length of 136 bp and differ at 7 positions (Tolun *et al.*, 1979; Shinagawa and Padmanabhan, 1980). Comparison of two Ad7 strains (Greider and Gomen) reveals that both repeats are 136 bp long but differ at 5 positions (Dijkema and Dekker, 1979; Shinagawa and Padmanabhan, 1980). Similar strain differences have also been found for Ad12. The length of the Ad12 ITR varies between 162 (Shinagawa and Padmanabhan, 1980) and 164 bp (Sugisaki *et al.*, 1980; Schwarz *et al.*, 1982). In all ITRs determined except one, a dCMP residue has been found at the 5' ends of adenovirus DNA. The exception is chick embryo lethal orphan (CELO) DNA, which has at its 5' end a dGMP residue (Aleström *et al.*, 1982a). In the ITRs of all human adenovirus DNAs, the sequence ATAATATACCTTAT (nucleotides 9–22) is present (Tolun *et al.*, 1979); the regions of the inverted repetitions beyond nucleotide 50 show a low degree of homology, although in all serotypes an asymmetrical distribution of base pairs is found. Comparison of the DNAs of the human serotypes with mouse strain FL DNA (Temple *et al.*, 1981) reveals that they have the sequence ATAATATAC (nucleotides 9–17) in common, while the homologous region between human adenovirus DNAs and CELO DNA is located between positions 9 and 15 (ATAATAT) (Aleström, *et al.*, 1982a). It is very likely that the conserved sequences 9–15 and 9–17 play a crucial role in the initiation of DNA replication and are probably involved in recognition of the site of initiation by the precursor of the terminal protein. In this respect, it is interesting to note that mouse

adenovirus strain FL DNA can be replicated in an *in vitro* DNA replication system of Ad2 DNA (Temple *et al.*, 1981). Shinagawa and Padmanabhan (1980) have pointed out that in Ad2, Ad3, Ad5, Ad7, and Ad12 DNA, an additional region of interesting homology is present. In these serotypes, the hexanucleotide TGACGT is found at or near the site where the sequences beyond the ITR begin to diverge. The function of this homology is unknown.

VI. TERMINAL PROTEIN

The presence of protein at the termini of adenovirus DNA was originally detected by Bellett and co-workers, employing DNA isolation procedures that avoid proteolytic digestion (Robinson *et al.*, 1973; Robinson and Bellett, 1975a). These investigators observed that the DNA-protein complex obtained is resistant to boiling and treatment with SDS, indicating that the protein is probably covalently linked to the DNA (Robinson *et al.*, 1973; Sharp *et al.*, 1976; Carusi, 1977; Padmanabhan and Padmanabhan, 1977).

When the buoyant densities of Ad2 and Ad5 DNA-protein complexes are compared with the densities of the corresponding DNAs isolated by digestion with pronase, a small difference of 2–10 mg/ml is found. This corresponds to an amount of protein present in the DNA-protein complex of a maximal 0.3% of the total virion protein (Robinson and Bellett, 1975a; Keegstra *et al.*, 1977). By gel electrophoresis of labeled DNA-free terminal protein (TP), it could be established that TP has an apparent molecular weight of 55K (Rekosh *et al.*, 1977).

Due to the hydrophobic character of TP, DNA-protein complexes aggregate very easily. As a result of this aggregation, DNA-protein complexes accumulate on tops of agarose and polyacrylamide gels during electrophoresis. It has been observed that when DNA-protein complexes are digested with restriction endonucleases and the digestion products are separated by gel electrophoresis, the terminal fragments carrying TP preferentially stay on top of the gel, while internal fragments conventionally run into the gel (Brown *et al.*, 1975; Sharp *et al.*, 1976). Another way to separate the DNA-protein complexes from protein-free DNA is based on differential binding of these compounds to glass-fiber filters (Coombs and Pearson, 1978; Coombs *et al.*, 1978).

To establish the nature of the DNA-protein linkage, deproteinized DNA and DNA-protein complexes have been subjected to enzymatic and nonenzymatic treatments. Both types of DNA are inaccessible to phosphatase, DNA polynucleotide kinase, and λ -exonuclease VII (Carusi, 1977; Sharp *et al.*, 1976), indicating that the 5' ends of adenovirus DNA are blocked. On the other hand, the 3' ends can freely be labeled with terminal transferase and are accessible to exonuclease III. These results are most easily explained assuming that in the DNA-protein complex,

TP is covalently attached to the 5' ends of the two complementary strands. The inaccessibility of deproteinized DNA is probably due to the fact that the 5' ends are still linked to short peptides. Treatment of DNA-protein complexes or deproteinized DNA with alkali or piperidine removes these peptides and makes the DNA freely accessible for enzymes (Robinson *et al.*, 1973; Carusi, 1977; Tolun *et al.*, 1979; Rekosh, 1981). TP can also be separated from adenovirus DNA by digestion with nuclease S1 (Ariga *et al.*, 1979; Roninson and Padmanabhan, 1980; Rijnders *et al.*, 1983). The DNA-protein complex is cleaved in close proximity to the protein-DNA linkage and yields a protein with a molecular weight of 55K (Rijnders *et al.*, 1983). Recently, Rekosh (1981) showed that treatment of the Ad2 DNA-protein complex with piperidine releases a protein with a molecular weight of 52K. This observation suggests that after DNase I or S1 digestion, the TP isolated still contains a few nucleotide residues.

The nature of the linkage between TP and the DNA molecule has been elucidated by Desiderio and Kelly (1981). Their experiments clearly indicate that Ad2 TP is bound to DNA by a phosphodiester bond between the hydroxyl group of a Ser residue of TP and the 5'-phosphate group of the terminal deoxycytidine residue of the two complementary strands of adenovirus DNA. The particular Ser residue in the TP amino acid sequence involved in the linkage of TP to DNA has recently been identified (Smart and Stillman, 1982).

The origin of TP has been uncertain for many years. Green *et al.* (1979c) showed by tryptic fingerprinting of TPs of five different human serotypes that these proteins were very similar in structure. On the other hand, Rekosh (1981) found different sizes for the TPs of different human serotypes, suggesting that TP is not of cellular origin. He concluded that TP is a highly conserved virus-coded protein. The viral origin of TP was unambiguously proved by Stillman *et al.* (1981), who showed that cell-free translation of mRNAs selected from a region between coordinates 11 and 31.5 on the viral l-strand (see Section IV) leads to synthesis of proteins with apparent molecular weights of 105, 87, and 75K. The 87K protein appeared to be identical to an 80K protein (Challberg *et al.*, 1980) that is covalently attached to the 5' ends of growing Ad2 DNA strands synthesized in an *in vitro* DNA replication system (Challberg and Kelly, 1979a,b). The 80K protein is structurally related to TP, suggesting that TP is synthesized as an 80K precursor TP (pTP) and that pTP is the active form of TP in adenovirus DNA replication. The different molecular weights found for pTP (80 and 87K) are due to the use of different molecular-weight markers. The 80/87K protein appears to be identical to the protein that is covalently attached to the DNA from temperature-sensitive (*ts*) mutant Ad2*ts1* virions grown at the nonpermissive temperature (Stillman *et al.*, 1981; Challberg and Kelly, 1981). Ad2*ts1* is a mutant that cannot cleave virus-coded precursor proteins to their mature counterparts during virion maturation (Bégin and Weber, 1975; Weber *et al.*, 1975).

The mapping of pTP on the virus genome led to the definition of a new early transcription unit, designated E2b. The structure of this region is discussed in detail in Section VII.B.3.

Evidence has been presented that TP plays an essential role in the initiation of adenovirus DNA replication. Analysis of the *in vitro* DNA replication system developed by Challberg and Kelly (1979a,b), in which the DNA-TP complex is used as a template, showed that the first step in the replication of adenovirus DNA is the linkage of dCMP to pTP. The protein probably recognizes a specific sequence within the inverted terminal repetition, which might be involved in binding of pTP to the DNA (Tamanoi and Stillman, 1982). It is likely that the conserved sequence 9-22 in different adenovirus serotypes functions as such a recognition sequence. The presence of TP in the DNA-TP complex might stabilize the initiation complex. Recently, it was shown that the protein is dispensable (Tamanoi and Stillman, 1982), since adenovirus DNA devoid of TP or remaining amino acids can also be used as template in an *in vitro* DNA replication system. It has been proposed that the presence of TP in the DNA-TP complex protects the viral DNA against nucleolytic degradation.

A protecting function of TP has also been proposed to explain the high infectivity of DNA-protein complexes. Deproteinized DNA is infectious when assayed by the calcium coprecipitation procedure (Nicolson and McAllister, 1972; Graham and van der Eb, 1973). However, the infectivity of DNA-TP complexes is 50-100 times higher (Sharp *et al.*, 1976; Chinnadurai *et al.*, 1978; van Wielink, 1978). Although the difference in infectivity might be due to a protective function of TP, it cannot be excluded that the presence of TP on the template is essential for accurate positioning of the pTP on the DNA during the first stage of initiation of adenovirus DNA replication. The role of TP in DNA replication is discussed extensively in Chapter 7.

VII. ORGANIZATION OF THE ADENOVIRUS GENOME

For the unraveling of the organization of the adenovirus genome, a great variety of techniques have been employed, i.e., DNA-RNA hybridization, R-loop mapping, genetic mapping of mutants, translation of preselected mRNA species, and nucleotide sequence analysis (for details, see Mautner *et al.*, 1975; Sambrook *et al.*, 1975; Grodzicker *et al.*, 1975, 1977; Chow *et al.*, 1977b, 1979a,b; Berk and Sharp, 1977a, 1978; Westphal *et al.*, 1976; Westphal and Lai, 1977; Kitchingman *et al.*, 1977; Kitchingman and Westphal, 1980; Miller *et al.*, 1980) (for sequences, see Appendix B). Despite a substantial nucleotide sequence divergence, all adenovirus serotypes studied so far show the general genetic organization (see Appendix B). Since the genomes of the highly homologous types Ad2 and Ad5 have been investigated most extensively, the organization of the

adenovirus genome is discussed employing for the most part data obtained with these particular serotypes. The precise location of major landmarks at the nucleotide level is indicated in the Ad2 sequence (Appendix B), unless otherwise stated. During the productive infection cycle of adenoviruses, the different viral genes are expressed in a rather complex pattern (Tooze, 1981; Persson and Philipson, 1982).

Traditionally, the adenovirus genes are subdivided into early genes, which are expressed before the onset of viral DNA replication, and late genes, which are transcribed after replication of adenovirus DNA has started. However, a group of intermediate genes has also been distinguished. These genes are expressed at intermediate times in infection in the absence of DNA synthesis and are also easily detected at late times. The complex transcription pattern of adenovirus DNA is discussed extensively in Chapter 5. A summary of the major RNA transcripts and the corresponding proteins is presented in Figs. 1 and 2. These diagrams demonstrate that the adenovirus genetic information is scattered over the

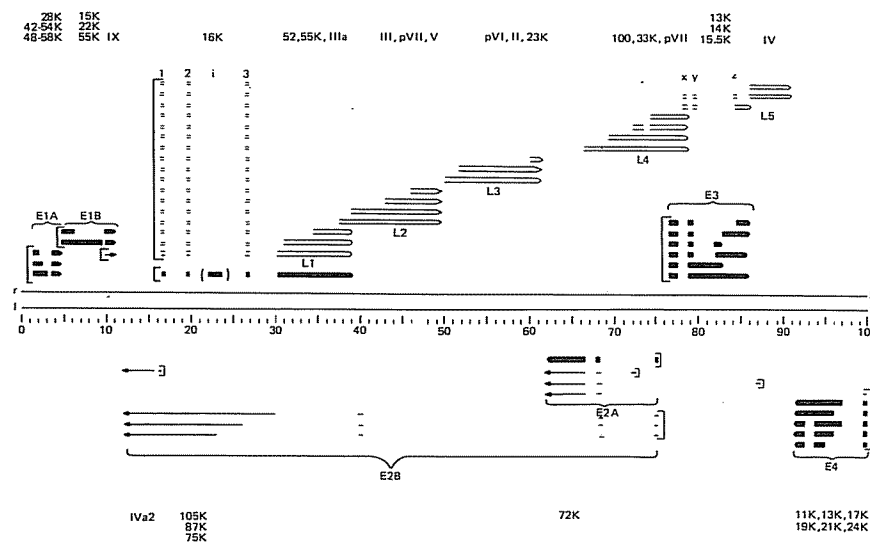


FIGURE 1. Transcriptional organization of the Ad2 genome. The genome is divided into 100 map units. The r-strand is rightward-transcribed into RNA and the l-strand leftward. The direction of transcription is indicated by arrows. The capped 5' ends of the cytoplasmic RNA indicate the positions of transcriptional promoters, while the arrowheads represent the 3' polyadenylation sites. Gaps in arrows indicate intervening sequences, which have been removed from the cytoplasmic RNA by splicing. The RNA shown in bold lines can be detected early in infection before the onset of DNA replication (regions E1a, E1b, E2a, E3, E4; also the late promoter at 16.5 units is active early in infection, leading to transcription to 39 units). The light lines represent intermediate RNAs synthesized at early as well as at late times in the infection cycle (E2a, E2b, polypeptide IX). The double-lined arrows indicate late RNA species. Correlations of mRNAs with encoded proteins are based on cell-free translation of selected RNA species and RNA mapping data. Proteins are designated by their molecular weights in kilodaltons (K) or by roman numerals (virion components).

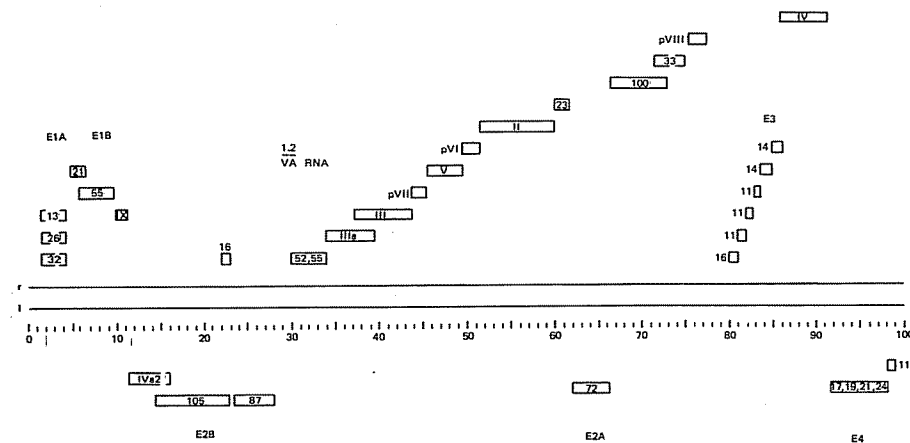


FIGURE 2. Protein-coding regions of the Ad2 genome. The regions on the adenovirus genome that code for protein have been determined by hybrid-arrest translation, by *in vitro* translation of preselected mRNAs, by RNA mapping, and by direct DNA and RNA sequence analysis. The identified proteins are designated by their apparent or theoretical molecular weights in kilodaltons or by roman numerals (virion components). Regions pVI, pVII, and pVIII indicate the positions of the precursors of polypeptides VI, VII, and VIII. Interrupted coding regions indicate discontinuous genes.

two complementary strands. About 69% of all genetic information is located on the rightward-transcribed strand (r-strand), while only 31% of the coding sequences are present on the leftward-transcribed strand (l-strand).

The positions of promoters and starts of transcription have been mapped via a variety of methods (Berk and Sharp, 1977b; Pettersson and Mathews, 1977; Spector *et al.*, 1978; Seghal *et al.*, 1979; Wilson *et al.*, 1979; Chow *et al.*, 1979a,b; Shaw and Ziff, 1980; Akusjärvi and Persson, 1981a; Stillman *et al.*, 1981). Many of the positions of promoters have been correlated with sequences generally indicated as TATA or Goldberg-Hogness boxes. These AT-rich sequences are considered to represent a constitutive part of promoter signals (see Chapter 5). The genes expressed early in infection are transcribed from six different promoters (r-strand: positions 1.3, 4.6, 16.5, and 76.6; l-strand: 75.1 and 99.1). The intermediate genes are transcribed from promoters located at positions 9.7 on the r-strand and 16.1 and 75.1 on the viral l-strand. The long late transcription unit uses the major late promoter at map position 16.5 on the viral r-strand. All primary transcription products of adenovirus DNA are processed in the nucleus before entering the cytoplasm. They are capped with ⁷m^eG5'pppN at the 5' end, and they are polyadenylated at the 3' end. With one exception (polypeptide IX mRNA), all primary transcription products are processed into families of related mRNAs that share common 5' and 3' ends, but differ by alternative splicing (early

regions E1a, E1b, E2a, E3, and E4, intermediate regions E2b and IVa₂, and late regions L1, L2, L3, L4, and L5). It should be noted that in fact, analysis of the late transcription unit of adenovirus led to the original discovery of the phenomenon of RNA splicing. A detailed analysis of the transcription of the adenovirus genome is presented in Chapter 5. The organization of the transcriptional units of the adenovirus genome will now be described systematically from left to right. Since the organization of the Ad2 and Ad5 genomes has been investigated most extensively, these genomes are used for illustration.

The positions of major landmarks of the transcription units are indicated in Figs. 3–6 and Appendix B in the r- and l-strand sequences. It should be borne in mind that sequences of the r-strand of DNA are equivalent to RNA transcribed from the l-strand and that sequences of the l-strand of the genome are equivalent to mRNA transcribed from the r-strand. Unfortunately, the entire nucleotide sequences of Ad2 and Ad5 are not yet available, only a number of noncontiguous regions having been sequenced. Therefore, the numbering of the base pairs in Fig. 3–6 and Appendix B has not been added, but the sequence of each specific region starts from the left with base pair number 1.

A. Early Region E1 (1.3–11.2)

Early region E1 is transcribed from the leftmost part of the viral r-strand. It contains genes involved in cell transformation (Graham *et al.*, 1974a,b; van der Eb *et al.*, 1979) and regulation of transcription (Berk *et al.*, 1979; Jones and Shenk, 1979a; Nevins, 1981). The complete nucleotide sequence of this region has been established for human serotypes Ad2, Ad5, Ad7, and Ad12 (van Ormondt *et al.*, 1978, 1980a,b; Sugisaka *et al.*, 1980; Dijkema *et al.*, 1980a,b, 1981; Bos *et al.*, 1981; Kimura *et al.*, 1981; Gingeras *et al.*, 1982). The overall organization of this region appears to be very similar for the different serotypes (van Ormondt *et al.*, 1980b; Dijkema *et al.*, 1982). The region between 1.3 and 11.2 m.u. can be subdivided into three transcription units designated E1a, E1b, and region IX (Kitchingman *et al.*, 1977; Berk and Sharp, 1977a, 1978; Chow *et al.*, 1979a,b). The mRNAs derived from region E1 have been characterized by EM mapping, *in vitro* translation, and sequence analysis. It appears that all mRNAs except protein IX mRNA have a spliced structure and code for a variety of proteins, some of which are structurally related.

1. Early Region E1a (1.3–4.6)

Early region E1a is transcribed from the r-strand between 1.3 and 4.6 m.u. and codes for proteins that are involved in initiation of transformation (van der Eb *et al.*, 1979) and regulation of early gene expression

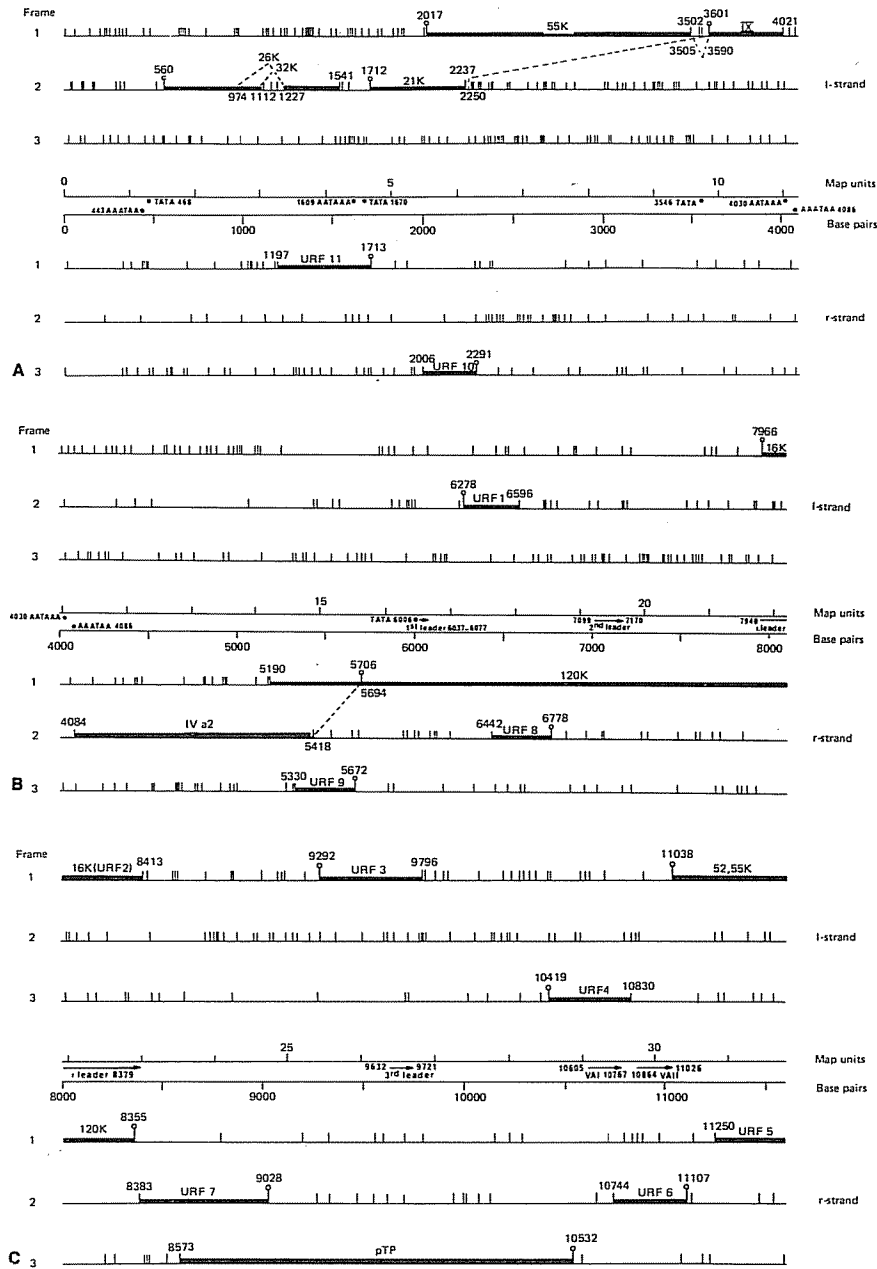


FIGURE 3A-C. Structural organization of the region between coordinates 0.0 and 31.7 on the Ad2 genome. The analysis of the structural organization is based on the nucleotide sequence shown in Fig. 18 [Appendix B], and indicated positions refer to this sequence. The l-strand of the DNA is homologous to r-strand transcripts, while the r-strand is homologous to l-strand transcripts. Here and in Figs. 4-6 and Appendix B: Termination codons (TAA, TGA, and TAG) are indicated in the three frames of the l- and r-strands by short vertical

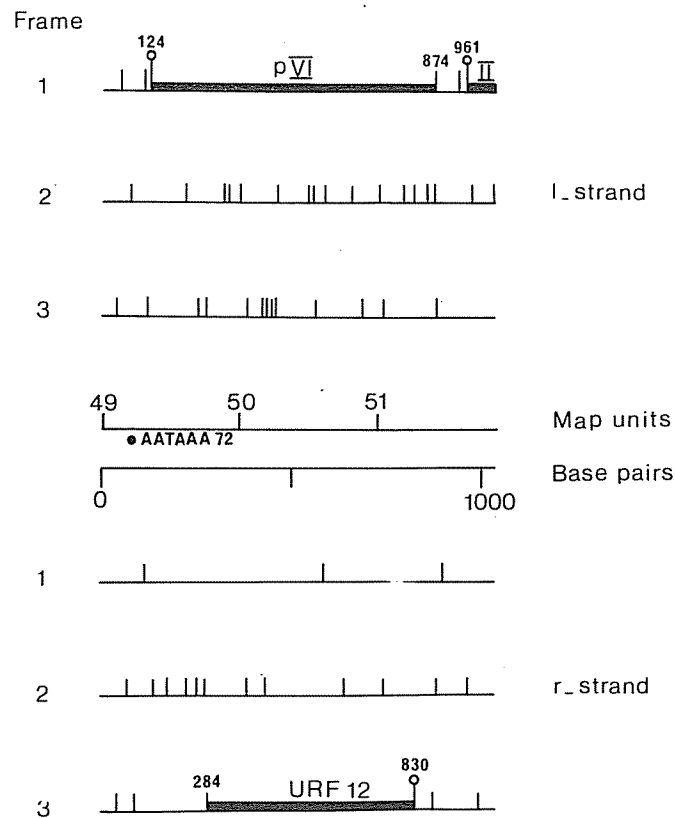
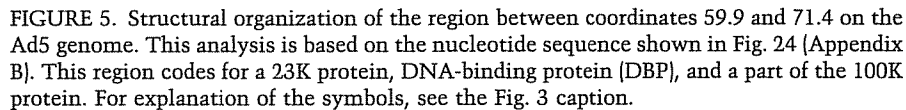


FIGURE 4. Structural organization of the region between coordinates 49.0 and 51.8 on the Ad2 genome. This analysis is based on the nucleotide sequence shown in Fig. 19 (Appendix B). This region mainly codes for the precursor of polypeptide VI. For explanation of the symbols, see the Fig. 3 caption.

(Jones and Shenk, 1979a; Berk *et al.*, 1979) (see Fig. 3). The promoter of this region has been mapped at position 1.3 (Wilson *et al.*, 1979). Analysis of the Ad2 sequence reveals that at position 468 [see Fig. 18 (Appendix B)], the TATA box TATTTATA is present. Baker and Ziff (1980, 1981) have characterized the position where transcription of the E1a RNA is initiated. They found that all mRNAs start with a capped dAMP residue

lines, while the initiation codon ATG is indicated by the symbol Φ . The coding regions that have been correlated with known proteins are shown by bold lines and are designated by molecular weights of the corresponding proteins or by roman numerals. Unidentified reading frames [(URF) initiating with ATG and terminating with one of the termination codons] or open reading frames [(ORF) regions between two termination codons] longer than 300 nucleotides are also indicated. Between the scales for Map units and Base pairs, the positions of TATA boxes, polyadenylation signals, and leader sequences are indicated. At some positions along the genome, splicing may occur. These positions are indicated by interrupted lines.



Since the reading frames in the E1a mRNAs are the same, the proteins derived from these mRNAs share their N-terminal and C-terminal segments and differ only in the number of intervening amino acids. From the DNA sequence, the complete amino acid sequences of the proteins specified by the 13 and 12 S mRNA species can be predicted. Both proteins must be rich in Pro and Glu residues and have theoretical molecular weights of 32 and 26K, respectively. The protein derived from the 9 S mRNA has an estimated molecular weight of 13K. These proteins have been correlated with proteins produced during cell-free translation of isolated mRNAs (Lewis *et al.*, 1976; Pettersson and Mathews, 1977; Harter and Lewis, 1978; Green *et al.*, 1979a; Esche *et al.*, 1980; Spector *et al.*, 1980a,b; van der Eb *et al.*, 1979; Lupker *et al.*, 1980). These translation products with apparent molecular weights of 48–58, 42–54, and 28K are structurally related, which is in agreement with the nucleotide sequence of this region. The discrepancy between the theoretical and apparent molecular weights probably reflects the extremely high Pro contents of these proteins, which lead to aberrant migration in gels.

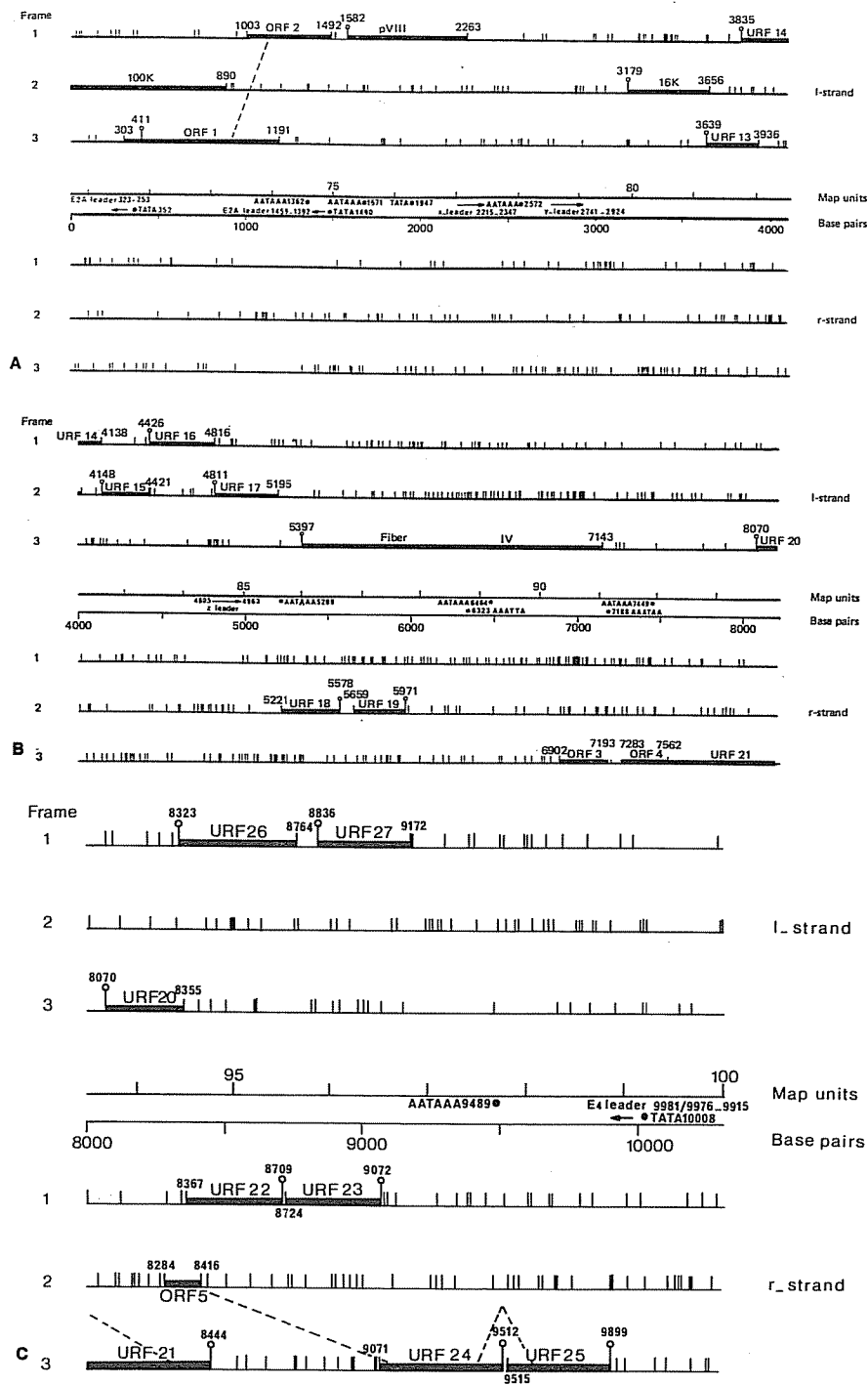


FIGURE 6A-C. Structural organization of the regions between coordinates 71.2 and 100.0 on the Ad2 genome. This analysis is based on the nucleotide sequence shown in Fig. 21 (Appendix B). For explanation of the symbols, see the Fig. 3-caption.

As mentioned before, the E1a regions of Ad2, Ad5, Ad7, and Ad12 show very similar organization. In all serotypes, three spliced mRNA species are synthesized. Recently, it was shown that the protein encoded by the 13 S mRNA governs early gene expression (Montell *et al.*, 1982).

2. Early Region E1b (4.6–11.2)

Early region E1b is transcribed from the viral r-strand between map coordinates 4.6 and 11.2 [see Figs. 3 and 18 (Appendix B)]. The proteins encoded by this region are involved in transformation and play an important role in oncogenesis; during lytic infection, these proteins are involved in DNA replication (Harrison *et al.*, 1977; Frost and Williams, 1978; Jones and Shenk, 1979a,b; van der Eb *et al.*, 1979; Bernards *et al.*, 1982; van den Elsen *et al.*, 1982). Little is known about the precise role of these proteins. Studies of cells transformed by DNA fragments of different length have suggested that region E1a is able to immortalize cells, while region E1b is required for full expression of the typical phenotype of adenovirus-transformed cells (van der Eb *et al.*, 1979; Houweling *et al.*, 1980).

The promoter of early region E1b is located at map position 4.6, where, at nucleotide 1670, a Goldberg–Hogness box TATATAA is found (Fig. 18). Transcription may start at position 1700 or 1702 (Baker and Ziff, 1981) and proceeds until nucleotide 4061 (Perricaudet *et al.*, 1980; Fraser *et al.*, 1982). The polyadenylation signal of region E1b is located at nucleotide 4030. The primary transcription product of region E1b is processed by splicing into a 22 and a 13 S mRNA species. Both species share a 3'-terminal segment from nucleotide 3590 to a polyadenylation site at nucleotide 4061. Both species also contain a 5'-terminal sequence from 1700 or 1702 to a donor splice site at nucleotide 2250. In the 13 S mRNA, nucleotide 2250 is joined to an acceptor splice site at 3590, whereas the 22 S mRNA includes nucleotide 2250 to a second donor splice site at nucleotide 3505. Nucleotide 3505 of the 22 S mRNA is ligated to the common acceptor splice site at nucleotide 3590. From these points, the mRNA sequence continues to the polyadenylation site near nucleotide 4061 (Perricaudet *et al.*, 1980; Aleström *et al.*, 1980). *In vitro* translation experiments have shown that two major proteins with molecular weights of 55–65 and 15–19K can be assigned to this transcription unit (Lewis *et al.*, 1976; Harter and Lewis, 1978; van der Eb *et al.*, 1979; Brackmann *et al.*, 1980). This observation is in agreement with the fact that the two mRNA species contain information for two major tumor (T) antigens with theoretical molecular weights of 21 and 55K, which are encoded by two overlapping reading frames. The 22 S mRNA codes for both proteins depending on which particular ATG triplet serves as the start codon. The 21K protein initiates at the 5'-proximal ATG (position 1712), while the 55K protein initiates at the second ATG (nucleotide 2017) in another reading frame (Anderson and Lewis, 1980; Bos *et al.*, 1981). In addition,

the 21K protein can also be synthesized from the 13 S mRNA. Peptide mapping has shown that the small-t and the large-T antigens do not share tryptic peptides, in accordance with the nucleic acid sequence data (Bos *et al.*, 1981).

Similar organization of region E1b has been found for Ad2, Ad7, and Ad12 (Bos *et al.*, 1981; Kimura *et al.*, 1981; Dijkema *et al.*, 1982; Gingeras *et al.*, 1982). This does not exclude small differences between mRNAs from different serotypes. Comparison of the E1b mRNAs of Ad5 and Ad12 has revealed that the Ad12 mRNA contains additional splices in the 3' noncoding part of the mRNA (Virtanen *et al.*, 1982a). The precise functions of the 21K and 55K proteins are still unknown.

The 22 and 13 S mRNAs both contain information for protein IX, a protein that has been mapped between 9.7 and 11.2 map units (Chow *et al.*, 1977b; Pettersson and Mathews, 1977; Esche *et al.*, 1980). However, this information is not translated from these messengers. Instead, a unique short mRNA is synthesized from an independent transcription unit between coordinates 9.7 and 11.2 (Wilson *et al.*, 1979; Chow *et al.*, 1977a,b; Pettersson and Mathews, 1977). The sequences of the genes that encode the Ad2 and Ad5 polypeptides IX have been established, which allowed the identification of transcription and translation signals (Maat *et al.*, 1980; Aleström *et al.*, 1980). The polypeptide IX TATA box is located at position 3546, and transcription starts at nucleotide position 3575 or 3577 (map position 9.7) in the Ad2 sequence [Fig. 18 (Appendix B)]. Its 3' end has been located at nucleotide position 4061 (map position 11.2) (Aleström *et al.*, 1980; Fraser *et al.*, 1982), while the polyadenylation signal AATAAA is located at position 4030. The same polyadenylation signal is also used for processing of the large and the small E1b T antigen mRNAs. The RNA synthesized is not processed and represents the only known unspliced adenovirus mRNA. The mRNA contains a continuous open reading frame that codes for a protein of 14K. Protein IX (apparent molecular weight 12.5K) is found in virions and was therefore originally classified as a late protein (Pettersson and Mathews, 1977). Later experiments showed that protein IX is also synthesized in the absence of viral DNA replication, indicating that it is an intermediate protein (Persson *et al.*, 1978). The complete nucleotide sequence of the polypeptide IX gene has been determined for human serotypes Ad2, Ad3, Ad5, Ad7, and Ad12 (Maat *et al.*, 1980; Aleström *et al.*, 1980; Dijkema *et al.*, 1981; Kimura *et al.*, 1981; Engler, 1981). Within the same group, the protein IX genes exhibit a striking similarity, but the genes of serotypes from different groups are much less homologous.

3. Unidentified Reading Frames

In the l-strand transcripts, a number of unidentified reading frames (URFs) have been detected. The URFs larger than 300 nucleotides are indicated in Figs. 3 and 18 (Appendix B). However, recently it could be

shown that in transformed cells and infected cells, an l-strand transcript is synthesized that spans the El_a–El_b junction and codes for a protein with a molecular weight of 11K [Katze *et al.*, personal communication]. This transcript might very well be derived from URF 11 located between nucleotides 1713 and 1197 on the viral l-strand. At position 443, the sequence AATAAA is found, which might function as a polyadenylation signal. This indicates that it is certainly not impossible that later some of these will appear to be expressed during the infection cycle, albeit at a very low frequency.

B. Late and Intermediate Genes in the Region between Coordinates 11.2 and 31

1. Major Late Promoter and Tripartite Leader

The region between 11.2 and 31 contains a mosaic of different strategic regions in both complementary strands [see (Figs. 3 and 18 (Appendix B))]. The major late promoter has been mapped on the r-strand at position 16.5 (Evans *et al.*, 1977; Ziff and Evans, 1978). This promoter is also active early in infection (Shaw and Ziff, 1980; Akusjärvi and Persson, 1981b). In the nucleotide sequence at this position, there is a TATA box TATAAA at nucleotide position 6006, and transcription starts from position 6037 (Baker and Ziff, 1981). During early times in infection, transcription proceeds no further than map position 39, while at late times, transcription proceeds to map position 99.0 (Fraser *et al.*, 1979). Messenger RNAs derived from r-strand transcripts starting at position 16.5 contain a common tripartite leader (Berget *et al.*, 1977, 1978; Chow *et al.*, 1977a,b; Akusjärvi and Pettersson, 1979a,b; Zain *et al.*, 1979a,b; Ziff and Evans, 1978). The sequence of the tripartite leader of late Ad2 RNA has been determined by sequencing complementary DNA (cDNA) transcribed from hexon mRNA and a cDNA clone of fiber mRNA (Zain *et al.*, 1979a; Akusjärvi and Pettersson, 1979b). The tripartite leader sequences have been established for a number of serotypes [Ad2 (Ziff and Evans, 1978; Akusjärvi and Pettersson, 1979a; Zain *et al.*, 1979a), Ad5 (van Beveren *et al.*, 1981), Ad3 and Ad7 (Engler *et al.*, 1981)].

The overall length of the Ad2 tripartite leader is 203 nucleotides, comprising 41 nucleotides from the promoter region at map position 16.5, 72 nucleotides from position 19.6, and 90 nucleotides from position 26.5 on the genome. Examination of the sequence reveals that the tripartite leader does not contain an AUG triplet, suggesting that translation of late adenoviral mRNA does not initiate within the tripartite leader. In some intermediate and late transcripts, an additional leader fragment (i-leader) has been detected by R-loop mapping, which maps at coordinates 21.5–23.0 (Chow *et al.*, 1979a). Sequence analysis has shown that in contrast to the tripartite leader, the i-leader (nucleotides 7940–8379) contains an open reading frame for a hypothetical protein of 15.9 kilodaltons (kd).

In vitro translation of mRNA selected on DNA fragments that contain i-leader sequences does indeed lead to synthesis of a hitherto unknown protein (URF2) with an apparent molecular weight of 13.6–16K (Lewis *et al.*, 1979; Lewis and Mathews, 1980; Virtanen *et al.*, 1982b). The termination codon for the 15.9-kd protein is not present in the i-leader, but is probably located within the third leader. The function of the 15.9-kd protein is still unknown.

2. Virus-Associated RNAs

At positions 28.8 and 29.5 on the genome, the genetic information for two low-molecular-weight RNAs is located, these RNAs being designated virus-associated (VA) RNAs VA-RNAI and VA-RNAII (Söderland *et al.*, 1976; Mathews and Pettersson, 1978) (Fig. 3). In contrast to all other genes, the VA genes are transcribed by RNA polymerase III instead of RNA polymerase II (Price and Penman, 1972; Weinman *et al.*, 1974, 1976; Söderland *et al.*, 1976). The VA-RNAs are probably synthesized from two separate promoter sites in the r-strand and do not undergo post-transcriptional processing. The genes and the RNA products have been subjected to nucleotide sequence analysis (Ohe and Weissman, 1970, 1971; Ohe, 1972; Pan *et al.*, 1977; Celma *et al.*, 1977a,b; Akusjärvi *et al.*, 1980). The nucleotide sequence of VA-RNAI was determined by Ohe and Weissman (1971) to be 157–160 nucleotides long (nucleotides 10,608–10,764/10,767). Vennström *et al.* (1978a,b) demonstrated that the 5' end of VA-RNAI is heterogeneous and may start at nucleotide 10,605 or 10,608 [Fig. 18 (Appendix B)]. The length of VA-RNAII is 158–163 nucleotides (nucleotides 10,864–11,021/11,026), and the two VA-RNAs are separated by a spacer about 98 nucleotides long. The function of these RNAs is still unknown; so far, no proteins derived from them have been found. It has been suggested that these RNAs play a role in splicing or stabilization of late mRNA (Murray and Holliday, 1979; Mathews, 1980). It is interesting to note that the VA-RNAs can form almost identical secondary structures with high stability. The structures show similarities to transfer RNA (Zain *et al.*, 1979b; Akusjärvi *et al.*, 1980).

3. Early Region E2b and Protein IVa₂ (11.2–30.2)

For a long time, it has been thought that the l-strand transcripts between map units 11 and 30 coded only for the intermediate protein IVa₂ (molecular weight 50K), a protein that is involved in the morphogenesis of virions (Persson *et al.*, 1979a). The gene of this protein has been mapped between coordinates 11.3 and 16.1 (Lewis *et al.*, 1975, 1977) [see Figs. 3 and 18 (Appendix B)]. Transcription of the IVa₂ gene starts from a promoter located at map position 16.1. Nucleotide sequences of this region reveal that although no regular TATA box is located in this region, the sequence TCCTT, which may resemble a TATA box, is pres-

ent at nucleotide 5859. RNA synthesis starts at position 5826 or 5824 and proceeds to nucleotide 4051 (Aleström *et al.*, 1980; Baker and Ziff, 1981; Fraser *et al.*, 1982) (Fig. 18). The messengers from this region contain an intron located between nucleotides 5419 and 5693 (Chow *et al.*, 1977a,b; Broker *et al.*, 1977; Kilpatrick *et al.*, 1979; van Beveren *et al.*, 1981). The mRNA contains a long open reading frame (ORF) corresponding to 445 amino acids of which the first 4 N-terminal amino acids are coded by RNA upstream from the donor splice site and the remaining amino acid residues by RNA downstream from the acceptor splice site. It is noteworthy that the reading frame in which these 4 N-terminal amino acids lie is part of a much longer reading frame that codes for a protein of 120 kd (see below). Another interesting feature of the IVa₂ gene is that the 3' end of the message overlaps the end of the E1b and polypeptide IX mRNAs with 9 nucleotides. Also, the IVa₂ termination codon TAA (nucleotide 4084) forms a part of the IVa₂ polyadenylation signal AATAAA (nucleotide 4086). The IVa₂ genes of serotypes Ad2, Ad5, and Ad7 have all been sequenced and show the same structural organization (van Beveren *et al.*, 1981; Engler and van Bree, 1982; Gingeras *et al.*, 1982; Aleström *et al.*, 1982b). The IVa₂ nucleotide sequences of Ad7 and Ad5 are 78% homologous.

A new class of mRNAs from the region between 11 and 30 m.u. was identified by Stillman *et al.* (1981). The promoter of these transcripts has been mapped at position 75.1 and is probably identical to the promoter of early region E2a. Transcripts of this region, which is designated E2b, contain, in addition to the 75.1-m.u. leader, additional leaders from 68.5 and 39 m.u. Region E2b has been classified as an intermediate transcription unit (Fig. 3). The main bodies of messages derived from this transcription unit may start at positions 30, 26, and 23, respectively, and continue to position 11.2. *In vitro* translation of preselected mRNAs derived from the region between 11.2 and 31.5 led to synthesis of proteins with molecular weights of 105, 87, and 75K (Stillman *et al.*, 1981; Binger *et al.*, 1982). The 87K protein is identical to the precursor terminal protein (pTP) with a molecular weight of 80K described by Challberg *et al.* (1980) (see Section VI). Nucleotide sequence analysis of this region has indicated the presence of two long ORFs located between 28.9 and 23.5 m.u. and 24.1 and 14.2 m.u. [Fig. 18 (Appendix B)]. The region between 28.9 and 23.5 m.u. beginning at nucleotide 10,577 has the first ATG at nucleotide 10,532 and continues to a terminator at nucleotide 8573. This frame codes for a protein with a minimum molecular weight of 74.5K. The second large ORF begins at nucleotide 8793, has the first ATG at 8355, and continues to a terminator TAG at nucleotide 5190. The total coding capacity of this reading frame is 132.1kd, while the capacity from the first ATG to the terminator is 120.4kd (Gingeras *et al.*, 1982; Aleström *et al.*, 1982; Engler *et al.*, 1983). Since the precise structure of the spliced E2b mRNAs is still unknown, it cannot be excluded that a part of the leader from map position 39 is part of the coding sequences of E2b mRNAs. EM

mapping of E2b mRNAs has indicated that the 3' ends of the messengers map at position 11.2, the same position where the 3' end of IVa₂ mRNA is located. It is therefore likely that the mRNAs of pTP and the 120kd polypeptide have the same 3' end and polyadenylation site as the IVa₂ mRNA (Aleström *et al.*, 1980; Stillman *et al.*, 1981). Smart and Stillman (1982) showed by analysis of tryptic peptides from the terminal protein and its precursor that the ORF between 28.9 and 23.5 codes for pTP. Very recently, the ORF from 24.1 to 14.2 was assigned to an adenovirus-specific DNA polymerase (Kelly, Stillman, and Hurwitz, personal communications). This polymerase has an apparent molecular weight of 140K, copurifies with pTP, and is able to complement a defective *in vitro* DNA replication system of the DNA-synthesis-negative temperature-sensitive (*ts*) mutant Ad5*ts*36 (Enomoto *et al.*, 1981; Lichy *et al.*, 1982; Kelly and Stillman, personal communications). The mutant Ad5*ts*36 has been mapped between 18.5 and 22.0 m.u. (Galos *et al.*, 1979). In addition to these two proteins, all E2b messengers contain genetic information for the IVa₂ protein, but this information is probably not translated from the E2b messengers.

4. Unidentified Reading Frames

Several unidentified shorter reading frames are present in this region of the viral genome (Fig. 3). However, no correlation with known proteins or gene functions has been discovered yet. In this respect, it should be noted that translation *in vitro* of early mRNA selected by hybridization to fragments of DNA derived from this region has identified mRNA species that encode additional proteins (Lewis and Mathews, 1980). A DNA fragment from 17.0 to 21.5 m.u. selects an mRNA that is complementary to the r-strand and codes for a 13.5-kd protein (Lewis *et al.*, 1979; Lewis and Mathews, 1980). Further, two polypeptides of 16.5 and 17.0kd have been described, translated from mRNAs that are selected by DNA fragments lying between 11.6 and 17.0 m.u. (Lewis *et al.*, 1979).

C. Late Regions L1, L2, and L3 (31.0–61.7)

A major event in the infection cycle of adenoviruses is the activation of the entire late transcription unit. As mentioned in Section VII.B.1, the promoter of the late transcription unit is located at map position 16.5, and this promoter is already active early in infection. However, during the early phase, transcription does not proceed further than map position 39 (Shaw and Ziff, 1980; Akusjärvi and Persson, 1981b). In the late phase, transcription continues to map position 99.0 (Fraser *et al.*, 1979, 1982). The transcription product ranging from map positions 16.5 to 99.0 is considerably processed, leading to the production of five families of late

mRNAs (L1–L5) (Chow *et al.*, 1977b; McGrogan and Raskas, 1978; Chow and Broker, 1978; Nevins and Darnell, 1978). Each of the five classes expresses more than one protein and contains mRNAs with a common 3' end (Ziff and Fraser, 1978; Nevins and Darnell, 1978; Fraser and Ziff, 1978). At the 5' end, all these mRNAs contain the tripartite leader.

The region on the Ad2 genome between 30.2 and 61.7 m.u. contains the genes for the families L1–L3. As mentioned above, the L1 family of RNAs is already expressed early in infection. This family consists of three mRNAs that have a common 3' end mapping at 39 m.u. At the same position, the polyadenylation site of the L1 family has been mapped (Fraser *et al.*, 1979, 1982). The L1 mRNAs code for two structurally related proteins of 52 and 55K (Lewis and Mathews, 1980; Miller *et al.*, 1980) and polypeptide IIIa (molecular weight 66K). Since nucleotide sequences from the left-hand end of Ad2 DNA have not been established further than position 31.5, only the initiation codon of the 52,55K protein has been identified unambiguously (Akusjärvi *et al.*, 1980). The function of the 52,55K protein is still unknown. The L1 family further contains genetic information for protein IIIa, which has been mapped by hybrid-arrest translation between 34.3 and 39.3 m.u. This protein has a molecular weight of 66K and is present in virions associated with the hexon polypeptides.

Located from positions 39 to 50 is the L2 family, consisting of three mRNA species that code for polypeptide III (molecular weight 85K), the precursor of polypeptide VII (20K), and polypeptide V (48.5K). These proteins are all constituents of adenovirus particles. One of these, the precursor of polypeptide VII, is processed during maturation of virions to mature polypeptide VII (molecular weight 18.5K). This protein is identical to the major core protein. The genes for protein III, the precursor of protein VII, and protein V have been mapped by R-loop mapping and hybrid-arrest translation at 37.4–43.9, 43.9–45.4, and 45.3–49.6, respectively (Miller *et al.*, 1980).

Fraser *et al.* (1982) have mapped the polyadenylation site of the L2 family at position 50. This fits well with the fact that in the nucleotide sequence from the region between coordinates 49.0 and 51.8 [Fig. 19 (Appendix B)], the polyadenylation site of the L2 family has been identified at nucleotide 92, while an AATAAA signal is present at nucleotide 72 (Akusjärvi and Persson, 1981a).

The nucleotide sequence data from region 49.0–51.8 make it possible to pinpoint exactly some landmarks of the L3 family of late mRNAs (see Figs. 4 and 19). Three species of mRNAs have been identified that can be translated into the precursor of polypeptide VI (pVI), hexon (polypeptide II), and a 23K protein. The gene for polypeptide pVI is located from 49.1 to 51.2 and has been sequenced completely (Miller *et al.*, 1980; Akusjärvi and Persson, 1981a). Also, the acceptor splice site at which the 5' leader sequences are joined to the pVI message has been determined (nucleotide 123) (Fig. 19). This splice site is situated very close to the

start codon (nucleotide 124). The gene for polypeptide pVI codes for a protein with a theoretical molecular weight of 27K. This protein is cleaved during maturation of young virions, resulting in the formation of polypeptide VI (molecular weight 24K), which is part of the adenovirion. With the help of nucleotide sequence analysis, the N-terminal end of the hexon polypeptide has been mapped at coordinate 51.6, while the C terminus is located at 59.7 (Akusjärvi and Pettersson, 1978a,b). The hexon polypeptide is translated from start codon 961 of an mRNA that contains, in addition to the tripartite leader, a main body starting at nucleotide 925 in the sequence of Fig. 19 (Appendix B) to nucleotide 836 in the sequence of Fig. 20.1. The common polyadenylation site of the L3 RNAs has been mapped at the same position. In accord with other polyadenylation sites, the sequence AATAAA is located close to this addition site (nucleotide 812) (Fig. 20.1). The total nucleotide sequence of the hexon gene has not been established yet; only stretches of nucleotides have been determined (Jörnvall *et al.*, 1981b). However, by combination of nucleotide sequence and amino acid sequence data, the complete amino acid sequence of the Ad2 hexon polypeptide has been established (Jörnvall *et al.*, 1981a). It appears that the hexon polypeptide of Ad2 consists of 966 amino acid residues. It is the largest viral protein and has a calculated molecular weight of 108K and an apparent molecular weight of 120K.

From positions 59.9 to 61.7, r-strand transcripts code for a protein of molecular weight 23K (Kruijer *et al.*, 1980; Akusjärvi *et al.*, 1981) [see Figs. 5 and 20.2 (Appendix B)]. A minor RNA species consisting of the tripartite leader and a main body corresponding to this region has been identified and translated. A protein with a molecular weight of 23K is synthesized from this messenger. Since the Ad2 mutant *ts1* has been mapped in the L3 region and is hampered in proteolytic cleavage of precursors of polypeptides VI, VII, and VIII, it has been suggested that the 23K protein is identical to a virus-coded protease (Bhatti and Weber, 1979).

D. Early Region E2a (61.5–75.1)

Early region E2a codes for the single-strand-specific, DNA-binding protein (DBP) (Figs. 5 and 6). This protein, discovered by van der Vliet and Levine (1973), is phosphorylated, has an apparent molecular weight of 72K, and is involved in DNA replication, in regulation of early and late gene expression, and in cell transformation (Ginsberg *et al.*, 1974; van der Vliet *et al.*, 1975, 1977; van der Vliet and Sussenbach, 1975; Carter and Ginsberg, 1976; Horwitz, 1978; Mayer and Ginsberg, 1977; Carter and Blanton, 1978; Nevins and Jensen-Winkler, 1980; Klessig and Grodzicker, 1979). The DBP genes of Ad2 and Ad5 have been analyzed in most detail. Therefore, the positions of strategic signals in the DBP gene are described in these sequences [Figs. 21 and 24 (Appendix B)]. It should be

pointed out that the main bodies of the Ad2 and Ad5 genes are highly homologous. The promoter for region E2a is located at 75.1 m.u. on the viral l-strand and is used early in infection (Baker *et al.*, 1979). At this position, the sequence TCCTTAA (nucleotide 1490) (Fig. 21) is found, which is an aberrant type of TATA box. This promoter is probably also used for transcription of the E2b transcription unit. At later times in infection, transcription of the E2a region starts from a promoter at map position 72.0, where the TATA box TACAAATTT is found (nucleotide 352) (Fig. 21). A minor start of transcription at intermediate and late times is found at 87 m.u. Recently, an additional promoter sequence was identified about 26 nucleotides upstream from the major early promoter (Mathis *et al.*, 1981). The function of the minor promoter sequence is still unknown.

Depending on the time in infection, mRNA species from the E2a region contain two different short leaders. Depending on the time post-infection, one is derived from position 75.1 [nucleotides 1392–1458/1459 (67/68 nucleotides long)] or 72.0 (nucleotides 253–321/323 (69/71 nucleotides long)] [Fig. 21 (Appendix B)]. The other is derived from position 68.8 [nucleotides 2936–3012 (77 nucleotides long)] (Baker *et al.*, 1979; Kruijer *et al.*, 1981, 1983) [Fig. 24 (Appendix B)]. The main body of the E2a mRNAs is located between map positions 66.5 and 61.5 [Fig. 24 (nucleotides 2309–642)] (Kruijer *et al.*, 1981; Akusjärvi *et al.*, 1981). The site of polyadenylation has been localized at nucleotide 642, while the sequence AATAAA is found at position 661 (Akusjärvi *et al.*, 1981; Fraser *et al.*, 1982). From the nucleotide sequence of the E2a region and the structure of DBP mRNAs, it can be derived that all coding sequences of these RNAs are located within the main body (Kruijer *et al.*, 1981, 1982). Translation starts at ATG 2300 and runs to stop codon 713. The Ad2 and Ad5 mRNAs code for a protein of 529 amino acids (molecular weight 59K), while Ad12 DBP is 484 amino acid residues long (molecular weight 54K). Comparison of the Ad2 and Ad5 DBP nucleotide sequences reveals a high degree of homology, with only 9 amino acid differences in the corresponding amino acid sequences. However, Ad5 and Ad12 DBPs differ considerably in nucleotide and amino acid sequences. These differences are mainly located in the N-terminal part of the DBP molecule. In contrast, the C-terminal regions of the DBP molecules show a high degree of homology (80%) (Kruijer *et al.*, 1983). It is especially this part of the molecule that is involved in DNA replication (Ariga *et al.*, 1980; Kruijer *et al.*, 1981). The terminal part of DBP is involved in regulation of late expression (Klessig and Grodzicker, 1979; Kruijer *et al.*, 1981).

E. Late Region L4 (66.5–77.3)

This region includes a set of r-strand transcripts that code for a 100-kd protein (66.5–73.1), a 33-kd protein (71.5–74.0), and the precursor of

polypeptide VIII (molecular weight 26K) (75.5–77.3) (Figs. 5 and 6). The indicated map positions have been determined by hybrid-arrest translation (Miller *et al.*, 1980). Polypeptide VIII (molecular weight 13K) is produced by proteolytic cleavage of its precursor during maturation of virions and is in virions associated with the hexon capsomers. The 100-kd protein is involved with folding of the hexon polypeptide chains into trimers (Ginsberg, personal communication), while the function of the 33-kd protein is still unknown. The four mRNAs that code for these proteins form the L4 family of late mRNAs and share the 3'-terminal sequences. The common polyadenylation site has been mapped at 78 map units.

Nucleotide sequences of this region have been determined in Ad2 and Ad5 DNA (Galibert *et al.*, 1979; Hérissé *et al.*, 1980; Kruijer *et al.*, 1981, 1982). Therefore, the strategic landmarks of the L4 proteins can be indicated at the nucleotide level. The acceptor splice point of the Ad5 100-kd polypeptide has been determined by reverse transcription of 100-kd mRNA and is located at nucleotide 2316 [Fig. 24 (Appendix B)] (Kruijer *et al.*, 1983). The polyadenylation site of the L4 mRNAs is mapped close to the sequence AATAAA at nucleotide 2572 [Fig. 21 (Appendix B)] (Fraser *et al.*, 1982). Comparison of the Ad5 sequence, which extends to coordinate 71.4, with the sequence of Ad2 indicates that nucleotides 3855–4107 of the Ad5 sequence (Fig. 24) are colinear with nucleotides 1–253 of the Ad2 sequence (Fig. 21). The frames in the overlapping sequences are identical and code, with a single exception, for identical amino acids. Using the combined sequences, it is possible to construct a hybrid 100-kd protein consisting of an amino-terminal part from Ad5 and a carboxy-terminal part of Ad2. The hypothetical hybrid protein consists of 805 amino acids and has an actual molecular weight of 89K.

The coding sequences of the 100 and 33-kd proteins partially overlap. However, since these proteins do not share tryptic peptides (Gambke and Deppert, 1981), it is most likely that they are encoded by r-strand transcripts in different ORFs. While the information for the 100-kd protein terminates at nucleotide 890, two ORFs (ORFs 1 and 2) can be distinguished in the other two reading frames, viz., ORF 1 from nucleotides 306 to 1191 (between stop codons 303 and 1191) and ORF 2 from nucleotides 1006 to 1492 (between stop codons 1003 and 1492 (Fig. 21). An ATG is present at nucleotide 411. Since one of the L4 mRNAs contains an internal splice that maps reasonably well in the region where these two ORFs overlap, it is likely that these regions code for the 33-kd protein. However, this has still to be proved by experimental data. One of the three short additional leaders for the fiber mRNA (x-leader) is also transcribed in this region from the r-strand (77.2–77.6). The x-leader has not been sequenced yet, but employing EM mapping data and typical RNA splice-site sequences, it has been inferred that this leader is transcribed from the r-strand from nucleotides 2215 to 2347. The l-strand between 66.5 and 77.3 units codes for the DBP mRNA leaders from positions 75.1,

72.0, and 68.8, respectively. The structure of the corresponding TATA boxes and individual leaders was described in Section VII.D.

F. Early Region E3 (76.6–86.0)

This region, located between coordinates 76.6 and 86.0, codes for a large number of r-strand transcripts and polypeptides (Fig. 6). At least six major species of mRNAs have been identified, coding for proteins of 13, 14, 15.5–16, and 19–21 kd, respectively (Lewis *et al.*, 1976; Harter *et al.*, 1976; Green *et al.*, 1979d; Ross *et al.*, 1980). The polypeptides of 19–21 kd are glycoproteins, which are associated with the membrane fraction (Persson *et al.*, 1979b, 1980a). Tryptic peptide analysis has shown that the 16-kd polypeptide is the unglycosylated precursor of the 19-kd protein (Persson *et al.*, 1980b).

The mRNAs from this region share sequences at their 5' ends from coordinates 76.6 to 77.6, which are ligated to sequences starting at 78.6 m.u. The 3' ends of the transcripts may vary.

Nucleotide sequence analysis of this region has revealed that a TATA box of the structure TATAA is located at nucleotide 1947 (76.7 m.u.), while transcription starts at nucleotide 1976/1978 (Baker and Ziff, 1981) [Fig. 21 (Appendix B)]. In region E3, two polyadenylation sites are present, one of which has been mapped at the nucleotide level (nucleotide 4148). Examination of the sequence of this region reveals that the sequence ATTAAA is found at position 4136. This sequence differs from the common hexanucleotide AATAAA that is found in all other Ad2 mRNAs associated with the polyadenylation site. In the sequence of region E3, the sequence AATAAA is located at nucleotide 5209, which fits very well with EM mapping data of some E3 mRNA species. However, for these messengers, the polyadenylation site has not yet been determined in detail.

The first ATG in the E3 region is found at position 2266, which suggests that E3 mRNAs have a 290-nucleotide-long untranslatable region at their 5' ends. About 80 nucleotides downstream from this ATG lies a potential splice site, and this site fits very well with the position where the common leader sequence of E3 mRNAs has been mapped (positions 76.6–77.6). This leader sequence may code for 27 amino acid residues, which would be common to all E3 proteins. However, determination of the amino-terminal sequence of the unglycosylated 16-kd protein has shown that translation of the coding sequence of this protein starts at nucleotide 3179 and continues to nucleotide 3656. This codes for a protein of 159 amino acids with a molecular weight of 18.4K. Obviously, the ATG at position 2266 present in all E3 mRNAs is not recognized during translation. If the 3' splice point of the first E3 intervening sequence is located around position 2840 (Hérissé *et al.*, 1980), this implies that the mRNA for the 16-kd protein has an untranslated region

more than 700 nucleotides long. Region E3 contains a number of short URFs. A hypothetical organization of translation is indicated in Fig. 6. Unfortunately, no data are available to assign the URFs unambiguously to individual proteins. As described above, the only exception is the 16-kd protein. The function of the E3 proteins is completely obscure. In some adenovirus-simian virus 40 hybrids, this region is absent without affecting the viability of the virus. Apparently this region is nonessential for viral multiplication (for a review, see Tooze, 1981). In addition to the E3 proteins, this region codes for two additional leaders of the fiber mRNAs, viz., the y-leader (78.6–79.2) and the z-leader (84.7–85.1) (Chow and Broker, 1978). Only the y-leader has been sequenced and appears to be located at nucleotides 2741–2924 (Zain *et al.*, 1979a). Employing EM mapping data and the common sequences of RNA splice sites, it has been inferred that the z-leader is located at nucleotides 4805–4963 (Hérissé *et al.*, 1980).

G. Late Region L5 (86.0–91.3)

The L5 family of late transcripts consists of two major mRNA species that code for a single virion protein, the fiber (polypéptide IV). The main bodies of these RNAs map between coordinates 86.0 and 91.3 (Miller *et al.*, 1980) (Fig. 6). RNA from this region differs from all other late messengers in that it may contain, in addition to the common tripartite leader, additional leader sequences (x, y, and z) from map positions 77.2, 78.6, and 84.7 (Chow and Broker, 1978; Zain *et al.*, 1979a). The y-leader is the most abundant additional leader of fiber mRNA; however, even this leader is not present in all RNA species. It has been shown that the presence or absence of the y-leader does not influence the translation of fiber mRNA. Even in the absence of the y-leader, the mRNA can be translated normally to fiber protein in an *in vitro* translation system (Dunn *et al.*, 1978). The nucleotide sequence of this leader has been established to be 184 nucleotides long, and although an ATG is present in this sequence, it is obviously not employed and not required for appropriate translation of fiber mRNA.

The complete nucleotide sequence of region L5 has been established (Zain *et al.*, 1979a; Zain and Roberts, 1979; Hérissé and Galibert, 1981; Hérissé *et al.*, 1981; Gingeras *et al.*, 1982) [Fig. 21 (Appendix B)]. The 5' end of the main body of the fiber mRNA is located at nucleotide 5395, adjacent to the codon of fiber mRNA at position 5397 (Zain and Roberts, 1979; Zain *et al.*, 1979a). The termination codon of the fiber gene is located at nucleotide 7143 and is part of the polyadenylation signal AA-TAAA at position 7141. The mRNA codes for 582 amino acid residues that constitute a protein with a theoretical molecular weight of 61.9K, which agrees very well with the apparent molecular weight of the fiber protein of 62K.

H. Early Region E4 (91.3–99.2)

Early region E4 messengers are transcribed from the viral l-strand between coordinates 91.3 and 99.0 and code for a large set of polypeptides (Fig. 6). The promoter of this region has been mapped at 99.2 m.u., while the 3' ends of E4 RNAs have been localized at 91.3 m.u. (Berk and Sharp, 1978; Chow *et al.*, 1979a,b; Baker and Ziff, 1981; Hashimoto *et al.*, 1981).

All E4 mRNAs share their 5'- and 3'-terminal nucleotide sequences, but vary in the location of splice points (Berk and sharp, 1978; Chow *et al.*, 1979a; Kitchingman and Westphal, 1980). These messengers code for a number of polypeptides with molecular weights of 11, 13, 17, 19, 21, and 24K (Lewis *et al.*, 1976; Green *et al.*, 1979d; Ross *et al.*, 1980). As yet, these proteins have not been assigned unambiguously to individual mRNA species. Only the position of the acidic 11K polypeptide has been correlated to a specific region in the nucleotide sequence of this region (Hérissé *et al.*, 1981).

Besides the fact that the synthesis of the E4 proteins starts about 2 hr after infection, reaches a maximum around 3 hr, and then declines, these proteins seem to be nonessential for DNA replication, and their role is at present unknown.

Recently, the complete Ad2 nucleotide sequence of this region has been established (Shinagawa *et al.*, 1980; Hérissé *et al.*, 1981; Gingeras *et al.*, 1982) [Fig. 21 (Appendix B)], while for Ad5, the region between 97 and 100 m.u. has been determined (Steenbergh and Sussenbach, 1979) [Fig. 25.1 (Appendix B)]. At nucleotide 10,008 in the Ad2 sequence, a TATA box with the structure TATATATA can be recognized as part of a promoter sequence. Transcription begins with the sequence TTTTTA at nucleotides 9981–9976, leading to a heterogeneous array of starts (Baker and Ziff, 1981) (Fig. 21). All major species of mRNAs contain a leader sequence starting at the cap sites and probably terminating at nucleotide 9915, where a potential 5' splice site is located. This leader sequence is devoid of ATG able to play a role in initiation of translation. Therefore, such a signal should be located in the body of the various mRNA species spliced to this leader sequence. At the other end of the sequence, transcription terminates close to an AATAAA sequence, which is located at position 7188. This is consistent with EM mapping data of E4 RNAs. It should be pointed out that transcription sometimes proceeds beyond this point to coordinate 61.5, leading to the production of a minor species of E2a mRNA (see Fig. 1).

The nucleotide sequence of the E4 region reveals that a large number of short URFs are present in all three reading frames.

Comparison of the nucleotide sequence and the mRNA mapping data indicates that there is a reasonably good correlation between the mapping data and potential donor and acceptor splice sites in the sequence. From the predicted structure of the various spliced mRNA species, a hypo-

thetical translation pattern has been proposed (Hérissé *et al.*, 1981; Gingeras *et al.*, 1982). However, only in the case of the acidic 11K protein could its coding region be deduced with reasonable certainty from the nucleotide sequence to be located in URF 23. Further nucleotide sequence analysis of mRNAs and translation of individual mRNA species is required to determine unambiguously the relationship between individual RNAs and the corresponding proteins.

I. Unidentified Reading Frames

In addition to the URFs of early region E4, an additional ORF with a coding capacity of 12kd (ORF 3) is found in the viral l-strand transcripts (Fig. 6). This region is located between stop codons at positions 7193 and 6902 and starts with AAA (7190) (Fig. 21). At nucleotide 7166, the first ATG codon is found, while at nucleotide 6323, even the sequence ATTAAA is present, which resembles an aberrant type of polyadenylation signal also present in early region E3. It should be noted that although the major E4 transcription termination site has been mapped at 91.3 m.u., Nevins *et al.* (1980) have calculated that transcription termination takes place at 88.4 m.u., which corresponds very well with the sequence ATTAAA at nucleotide 6323 (Hérissé *et al.*, 1981). However, no mRNA species derived from this region are currently known. The same holds for two URFs in r-strand transcripts that code for proteins with theoretical molecular weights of 10.6 and 12K (URFs 26 and 27).

VIII. COMPARISON OF GENOMES AND CONCLUDING REMARKS

The organization of the adenovirus genome as described in Section VII has mainly been restricted to Ad2 because the most detailed information is available for this serotype. However, it should be emphasized that for all serotypes the structure of which has been investigated, the same overall organization has been observed. For a number of serotypes, nucleotide sequence data are available. These data are compiled in Appendix B, including the analysis of these sequences. For a number of genes, the nucleotide sequences have been compared, as well as the amino acid sequences of the corresponding proteins. Van Ormondt *et al.* (1980b) have analyzed the homology among the E1a regions of Ad5, Ad7, and Ad12, while Bos *et al.* (1981) and Kimura *et al.* (1981) have studied the homology of the E1b regions of Ad5 and Ad12. The IVa₂ and polypeptide IX genes of Ad2, Ad3, Ad5, and Ad7 have been compared (Dijkema *et al.*, 1981; Engler, 1981; Engler and van Bree, 1982), as well as the late leaders of Ad2, Ad3, and Ad7 (Engler *et al.*, 1981) and the E2b regions of Ad2 and Ad7 (Engler *et al.*, 1983). The redundancies of different serotypes were

analyzed by Tolun *et al.* (1979) and Shinagawa and Padmanabhan (1980), while the DNA-binding protein genes of Ad2, Ad5, and Ad12 were compared by Kruijer *et al.* (1981, 1982, 1983).

Detailed analysis of the organization of the adenovirus genome reveals that the available coding information of this virus is used in a very economical fashion. Unraveling of the information at the nucleotide level reveals all kinds of peculiar properties in its organization. There are spliced and unspliced mRNA species (e.g., hexon and polypeptide IX RNA), overlapping termination codons and AATAAA signals (e.g., fiber and IVa₂ RNA), overlapping genes (e.g., the 33- and 100-kd proteins), and symmetrical transcription (120-kd protein and the 16-kd i-leader product). There are classic TATA boxes (e.g., E1a proteins) and polyadenylation signals (AATAAA) (hexon RNA) and aberrant sequences with the same function [TATA box TCCTT (E2a early promoter) and polyadenylation signal ATTAAA (region E3)].

In conclusion, the adenovirus genome is a microuniverse in itself, and the study of its organization and regulation of expression is a great joy and satisfaction for every scientist who dedicates herself or himself to the unraveling of its secrets.

ACKNOWLEDGMENTS. The author gratefully acknowledges the very valuable assistance of Mr. O. van Hien for providing computer facilities and Dr. T. Broker for maps and other information. Without their help, this chapter would never have been completed. He also thanks M. M. Kwant, M. G. ter Braak-Kuijk, W. van Driel, F. M. A. van Schaik, E. Simon, W. Kruijer, A. W. M. Rijnders, J. van der Rijst, and H. Laanen for technical assistance and Dr. P. C. van der Vliet for critical reading of the manuscript. He gratefully acknowledges the fact that his colleagues Drs. J. Engler, R. J. Roberts, K. Fujinaga, M. Horwitz, U. Pettersson, H. van Ormondt, R. Padmanabhan, B. Stillman, E. Ziff, and F. Galibert have made available new data prior to publication.

APPENDIX A: RESTRICTION ENDONUCLEASE CLEAVAGE MAPS

This appendix contains a compilation of restriction maps of the genomes of different adenovirus serotypes (Figs. 7–17). These maps have partially been published and partially been presented as personal communications. Most of these maps have been compiled before by Tooze (1981) and are redrawn with permission from the Cold Spring Harbor Laboratory Publication Department. The coordinates of the Ad1, Ad2, and Ad5 maps have been recalculated (Gingeras *et al.*, 1982). Details on the origin of the maps are indicated in Tooze (1981), unless otherwise stated.

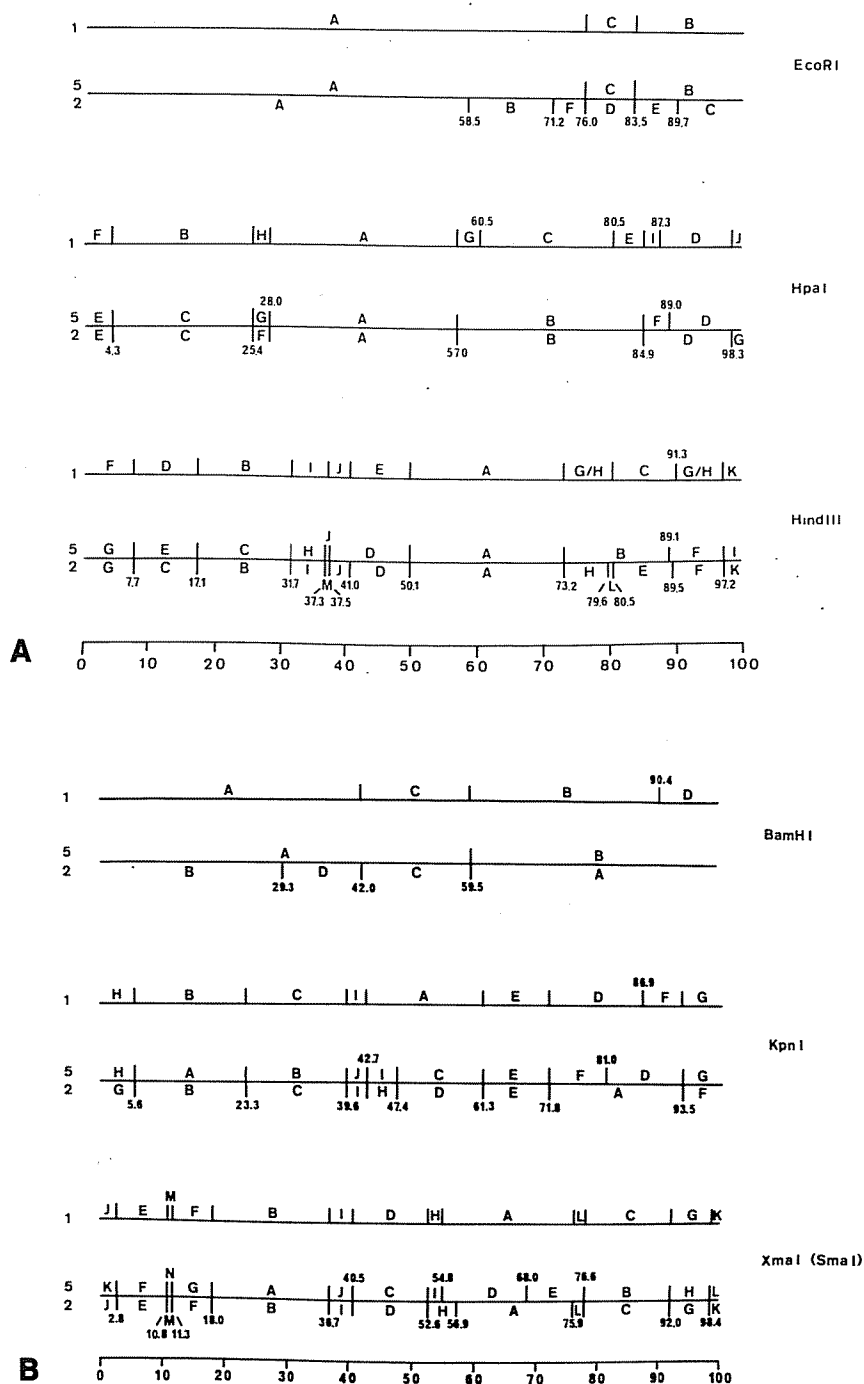


FIGURE 7A-D. Restriction endonuclease cleavage maps of Group C Ad1, Ad2, and Ad5.

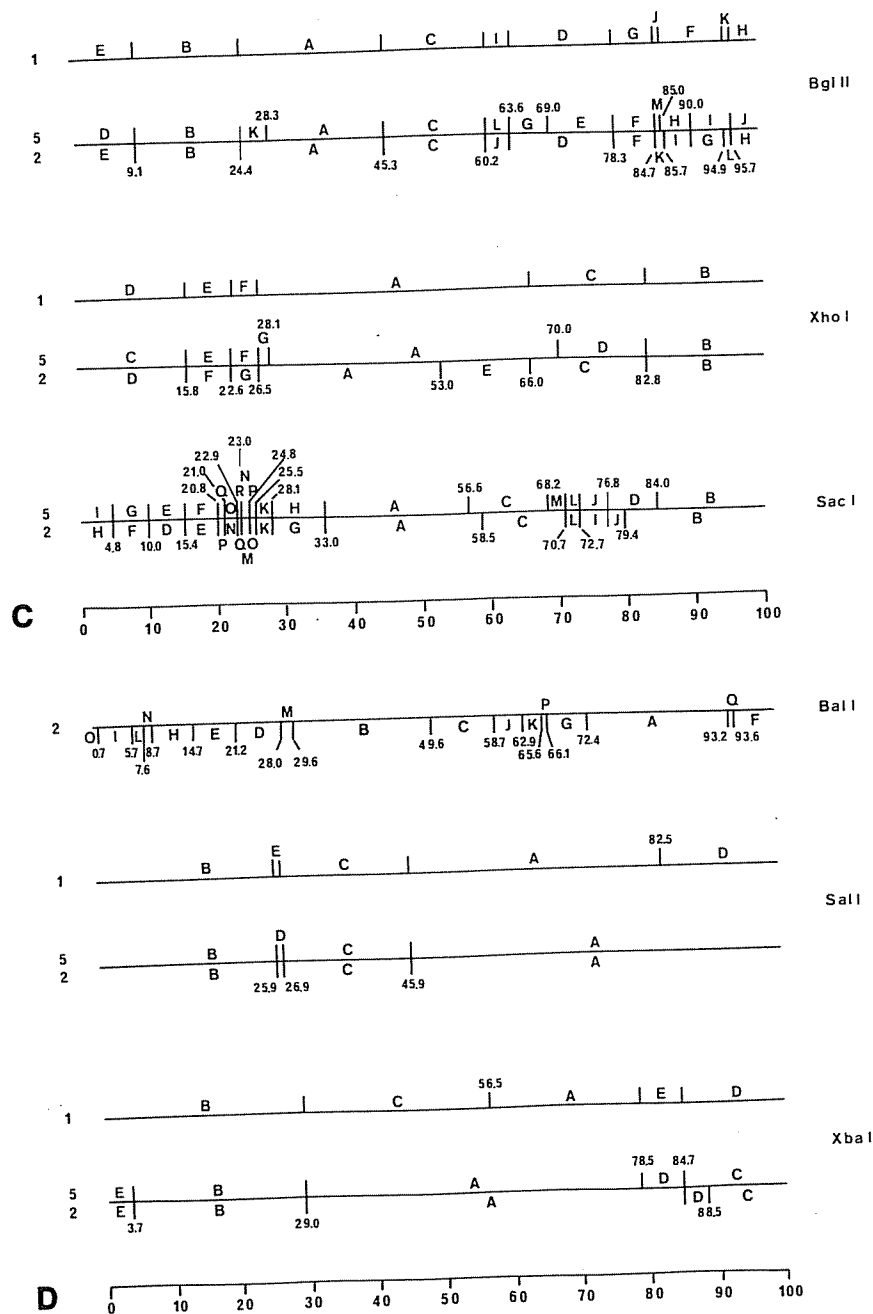


FIGURE 7 (Continued)

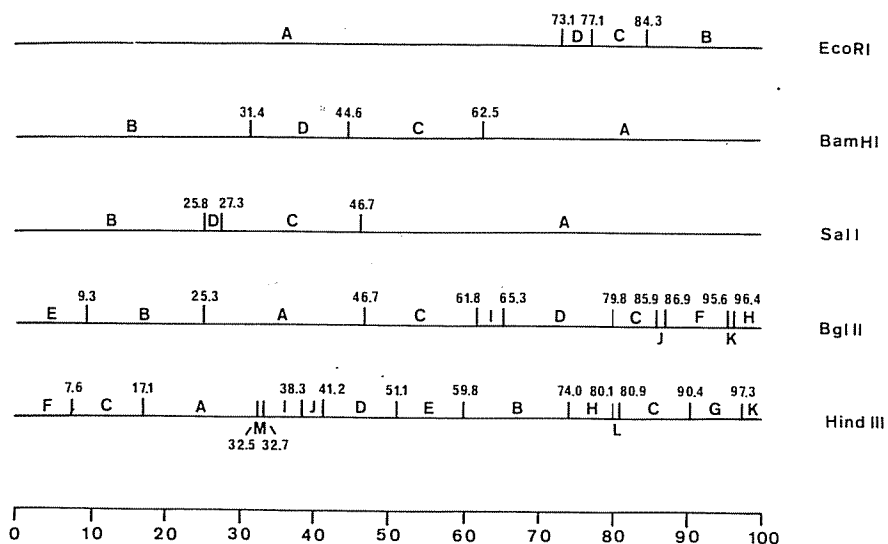


FIGURE 8. Restriction endonuclease cleavage maps of Group C Ad6. The maps were determined by Naroditsky *et al.* (1980) and oriented such that the transforming region is located at the left. The *EcoRI* map was determined by Forsblom *et al.* (1976).

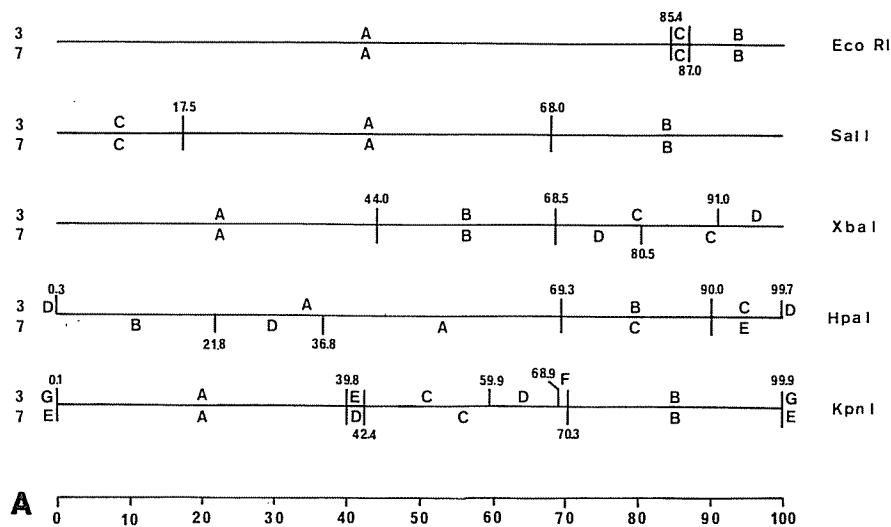


FIGURE 9A, B. Restriction endonuclease cleavage maps of Group B Ad3 and Ad7. The *BstEII* and *BclII* maps were determined by R. Padmanabhan (personal communication).

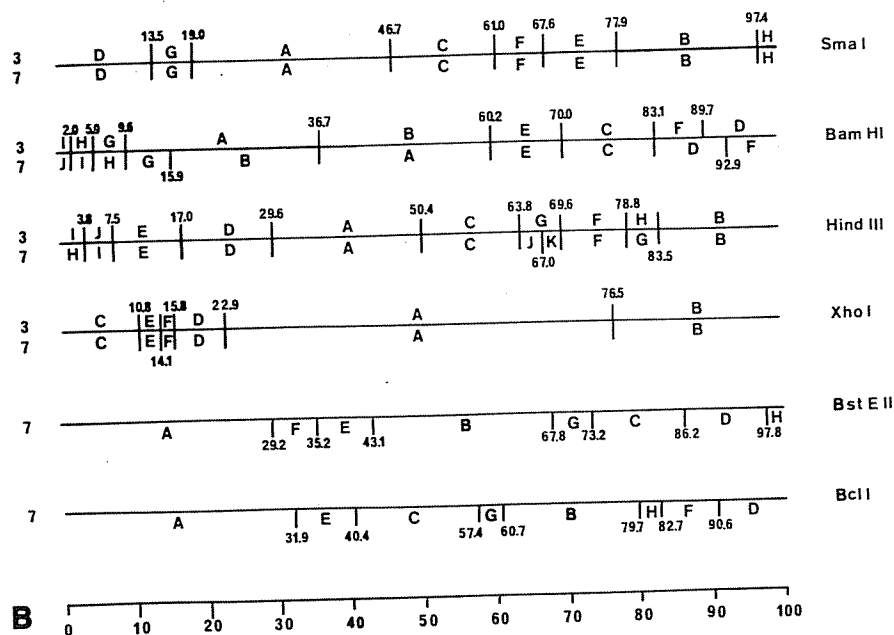


FIGURE 9 (Continued)

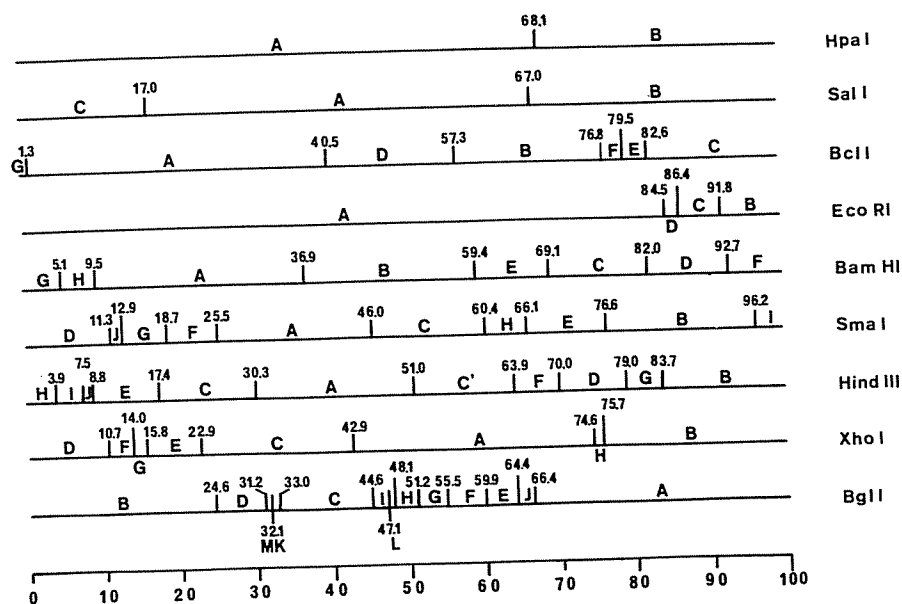


FIGURE 10. Restriction endonuclease cleavage maps of Group B Ad16 (Chang 79). These maps were determined by Varsanyi *et al.* (1977), Winberg and Hammarskjöld (1980), and Hammarskjöld and Winberg (personal communication).

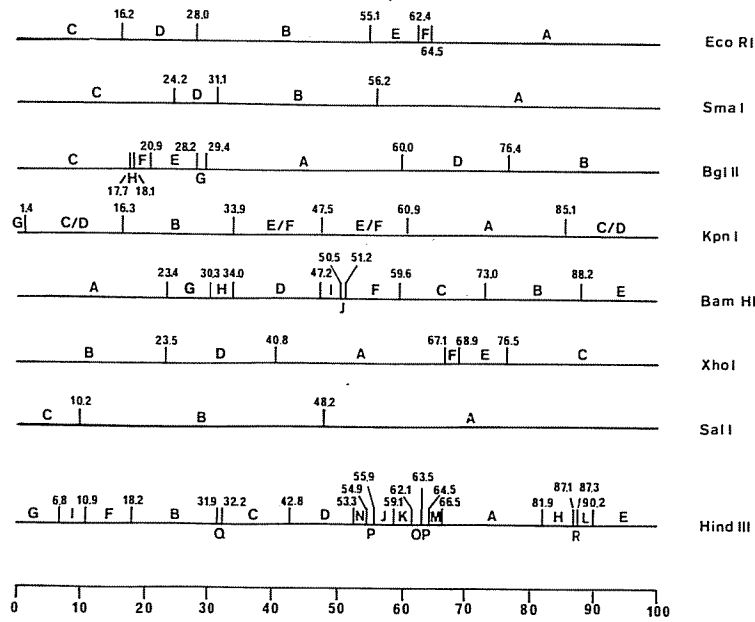


FIGURE 11. Restriction endonuclease cleavage maps of Group A Ad12 (Huie).

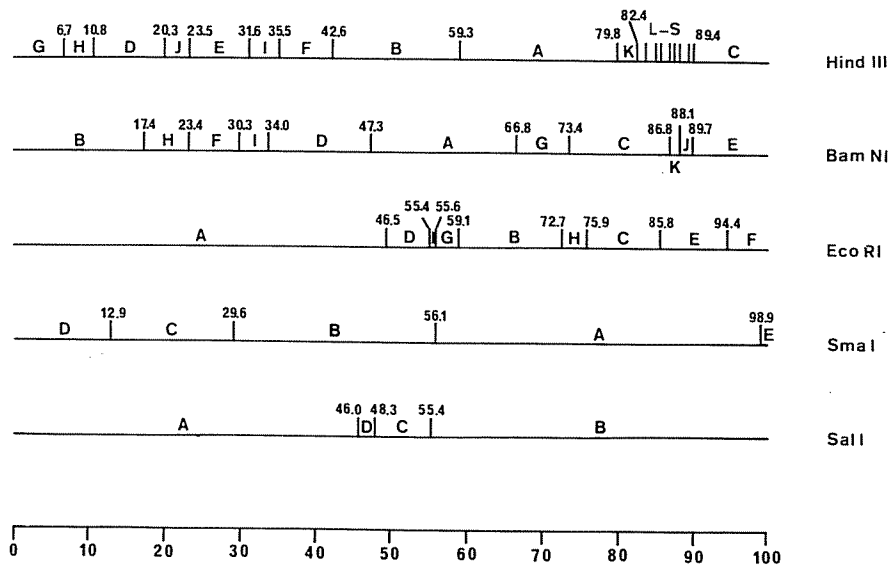


FIGURE 12. Restriction endonuclease cleavage maps of Group A Ad31 (strain 1315). The maps were determined by Y. Sawada, Y. Yamashita, F. Kamda, K. Sekikawa, and K. Fujinaga [personal communication].

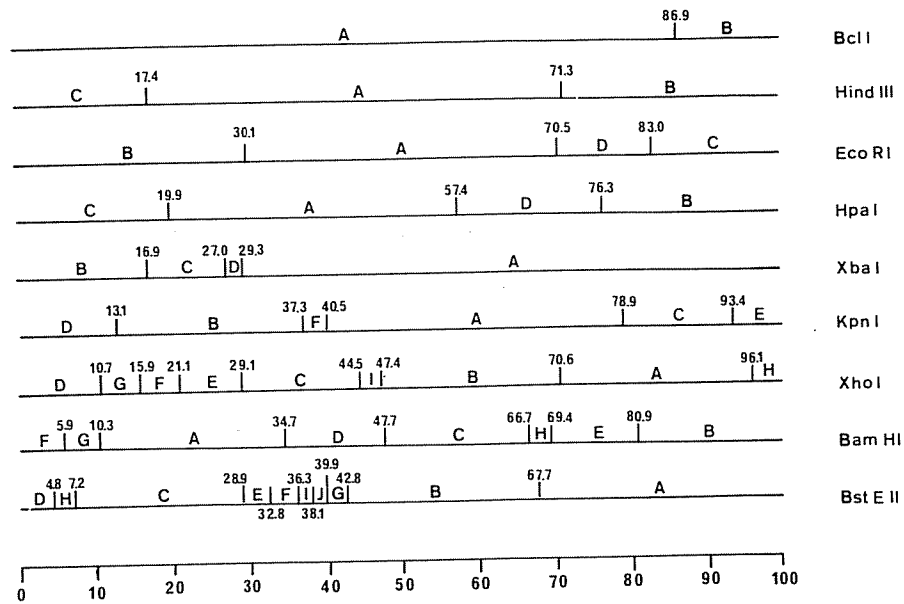


FIGURE 13. Restriction endonuclease cleavage maps of Group E Ad4. These maps were determined by Tokunaga *et al.* (1982).

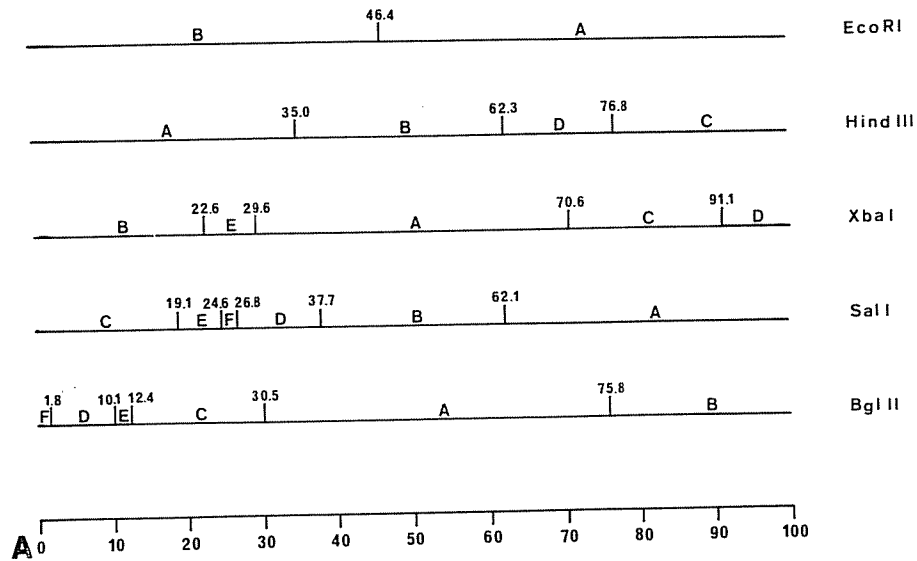


FIGURE 14A, B. Restriction endonuclease cleavage maps of simian adenovirus type 7. The *EcoRI*, *SalI*, and *BglII* maps of simian adenovirus (strain C8) were determined by Naroditsky *et al.* (1980) and oriented with respect to the conventional genetic map by Ponomareva *et al.* (1979), who located the transforming region to the left. The other maps were determined by T. I. Tikchonenko and colleagues (personal communication).

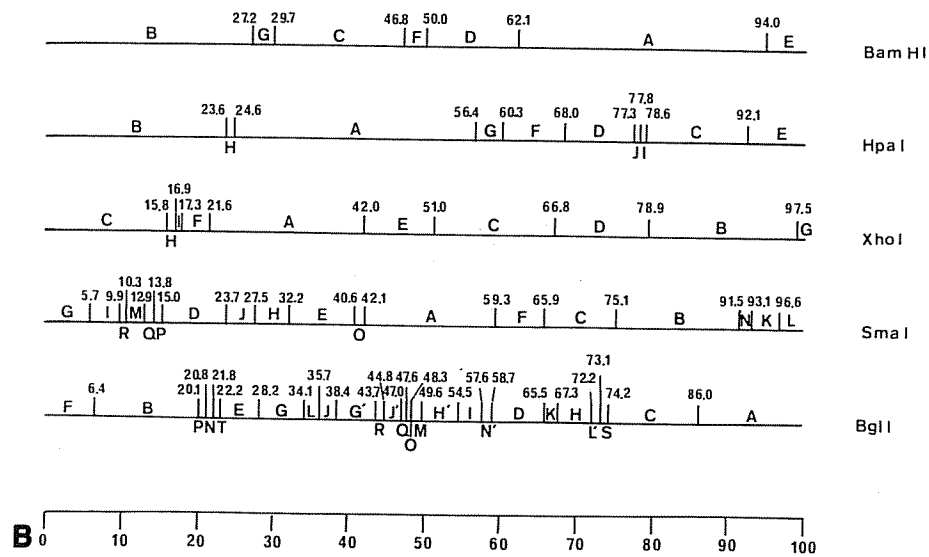


FIGURE 14 (Continued)

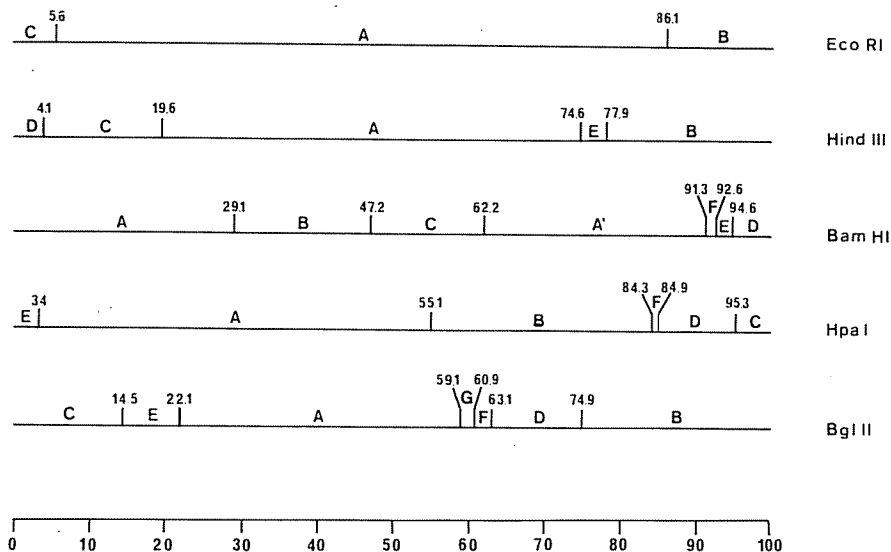


FIGURE 15. Restriction endonuclease cleavage maps of simian adenovirus type 20. These maps were determined by T. I. Tikchonenko and colleagues [personal communication].

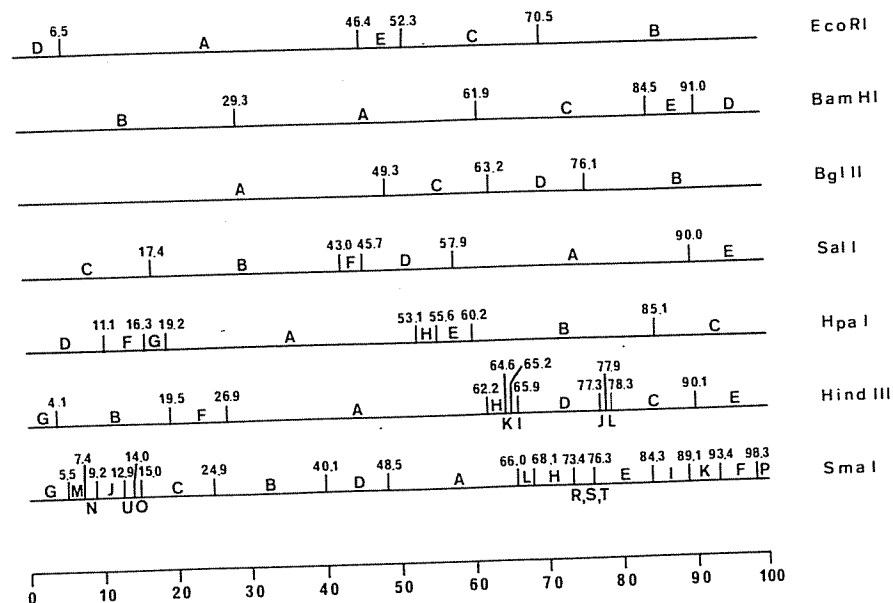


FIGURE 16. Restriction endonuclease cleavage maps of simian adenovirus type 30. The *EcoRI* and *BglII* maps were determined by Dimitrov *et al.* (1979). They were originally reported to be those of simian adenovirus type 38, and identification subsequently revised by Tikchonenko and colleagues (personal communication), who also determined the other maps.

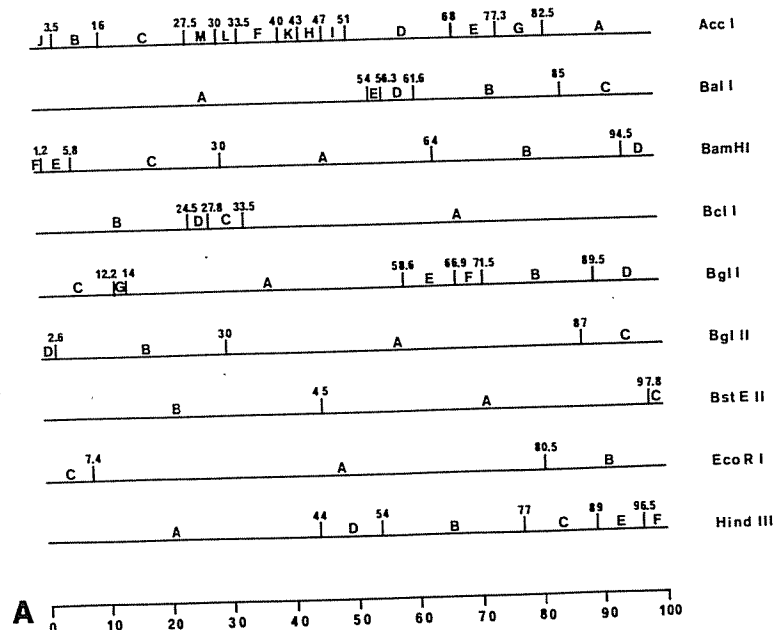


FIGURE 17A, B. Restriction endonuclease cleavage maps of mouse adenovirus type FL. These maps were determined by Larsen *et al.* (1979). For the orientation, see Larsen *et al.* (1979) and Temple *et al.* (1981).

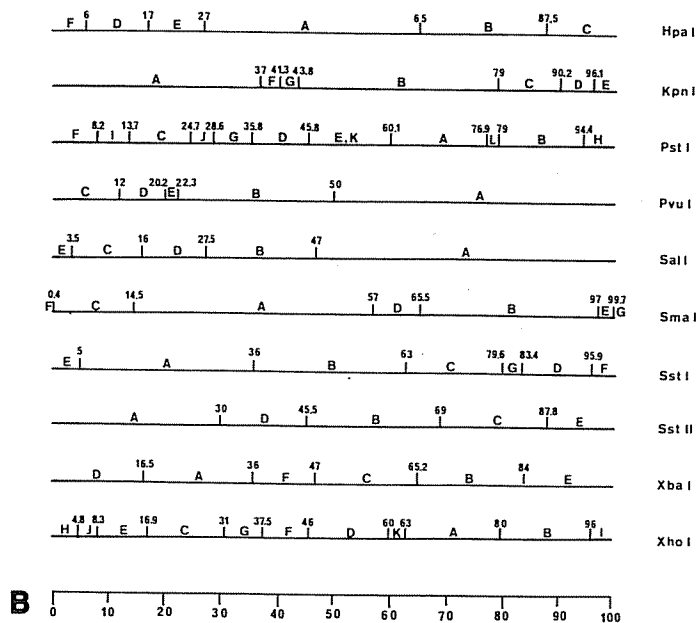
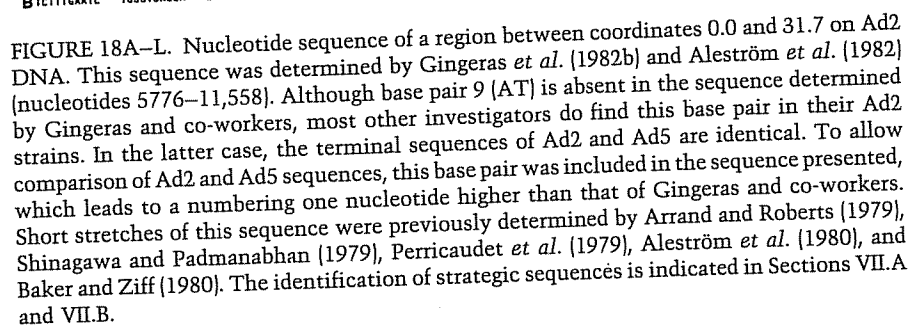


FIGURE 17 (Continued)

APPENDIX B: NUCLEOTIDE SEQUENCES

This appendix contains a compilation of nucleotide sequences partially published and partially presented as personal communications (Figs. 18–29). Since r-strand transcripts are homologous to the l-strand, the positions of important landmarks for r-strand transcripts are indicated in the l-strand sequence. Likewise, strategic sequences for l-strand transcripts are indicated in the r-strand sequence. The sequences of Ad2 and Ad5 are very homologous. Therefore, it has been supposed that specific signals identified in the sequence of one serotype also indicate the positions of these signals in the sequence of the other serotype. The positions of the inverted terminal repetition boundaries and start and termination codons of known coding regions are indicated, as well as the positions of 5' and 3' ends of mRNAs, splice points, and TATA boxes. The latter signals are supposed to be a constitutive part of transcriptional promoters. The sequences AATAAA and ATTAATA, which are found within about 30 nucleotides from the 3' end of the mRNAs, are underlined. These sequences have been associated with polyadenylation. Open reading frames (ORFs), defined as regions between two termination codons in the same frame, have been indicated when the size exceeds 300 nucleotides. The same holds for unidentified reading frames (URFs) (regions that start with an ATG codon and terminate with one of the termination codons).



stop URF 10	2020	2030	2040	2050	2060	2070	2080	2090	2100
CAAAATATTT	CCTATTATCC	TCGCTCTTT	GGGTAGACTC	GCCCCCATG	GACBACCTAA	AAGACCGGA	CGTAGACACC	TCTGCCACCC	ACTCTGTGTT
GTTTATAAA	GGATAATGG	AGCGAAGAA	CCCATCTGAG	CGGGGGGTAC	CTGCTGGATT	TTCGTGECAT	GCATCTGTGG	AGAGCGGTGG	AGAGCGGTGG
	start 55 kD protein								
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
GTTAGCGGAC	GATGACACAA	GAAGCGACCG	GGGCGCTTAT	TATGGCTGCC	TCCTGCTTGT	CGTCCCTCTT	CGGTCCCGCC	CGCCCGCGGT	CGTCCCTCTC
GAATCGCGTG	CTATGCTTGT	CTTCCGTCTG	CCCGGCAATA	ATACCGACCG	AGGAGCAACA	GCAGGAGGAA	GCCAGCGCGC	GCGCGCGGCA	GGAGCGAGAC
									start URF 10
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
GGTACCTTGG	GCTCTCGGCC	GGACCTGGGA	GGCCTTACTT	ACAACATGTC	CACCGACTTG	ACAAGGTCT	TGACTCTGGC	TAAATTTGTT	TATTGCTCTC
CCATGGAAAC	CGAGAGCGGG	CGTGACCCCT	CGGGAAATGA	TGTTGTACAG	GTGGCTGAAC	TGTTTCCAGA	ACTGAGACGC	ATTTTAAACA	TTAACGAGGA
			stop 21 kD protein	splice 13 S E18 RNA					
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
ACCGGTCCCTC	GATTTCCTCC	ATTTCTCCCT	CGCCCCCGGA	AGACTCCGAT	GTCTCTCCCG	ATCCTTAGAT	TGAAATTCGA	ATTACTGGTC	TGTGGGAGGA
TGGCGAGGGG	CTAAAGGGGG	TAAAGAGGGA	CGCGGGGGCT	TCTGAGGCTA	CAGAGAGGCG	TAGGAATCTA	ACTTTTAGCT	TAATGACCAG	ACACCGTCTC
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
CTCACACAAT	GAAGAAGTCG	CTAATTCTTA	TAAACGCGAT	TACTCGAAGT	AGACGACCGC	GTCTTCATAA	GGTATCTCGT	CGACTGGTGA	ATGACCGACG
GAGTGTGTGA	CTTTTACGCA	GATTAAAGTA	AATTGCGCTA	ATGAGCTTGA	TCGCTGGCGG	CAGGAATATT	CCATAGAGCA	GCTGACCACT	TACTGGCTGC
2510	2520	2530	2540	2550	2560	2570	2580	2590	2600
TGGGTCCCTC	ACTAAAGTCG	CTCCGATAAT	CCCATATATG	TTTCCACCGT	GAATCCGGTC	TAACTTTCAT	GTCTTAATCC	TTTGAATATT	TATAGTCTTT
AGCCAGGGGA	TGATTTTGGG	GAGGCTATTA	GGGTATATGC	AAAGGTGGCA	CTTAGGCGCA	ATTGCAAGTA	CAAGATTAGC	AAACTGTGAA	ATATCAAGAA
2610	2620	2630	2640	2650	2660	2670	2680	2690	2700
AACAACATTC	TAAAGACCGT	TGCCCGGGCT	CCACCTCTAT	CTATGCTCTC	TATCCGACCG	GAATCTACAA	TCGTACTATT	TATACACCGG	CCCCACAGAA
TTGTGCTTAC	ATTCTGCGGA	ACGGGGCGCA	GGTGAGAGTA	GATACGGAGG	ATAGGTGGCG	CTTTAGATGT	AGCATGATAA	ATATGTGGCC	GGGGGTGCTT
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
TCGTACTCTG	CCCAACAATA	ATACCTACAC	TCCAAATGAC	CAGGGTTAAA	ATCGCCATGC	CAAAAGGACC	GGTTATGGTT	AGAATAGGAT	GTGCCACATT
GGCATGGAGG	GGGTGGTTAT	TATGAATGTG	AGGTTTACTG	GTCCCAATTT	TAGCGGTACG	GTTTTCTCGG	CCAATACCAA	TCTTATCCTA	CACGGTGTAA
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
CGAAGATACC	CAAAATGTGA	TGGACACACC	TTCGGACCTC	GCTACATCTC	CAAGCCCGCA	CACGGAAAAA	GACBACGACC	TTCCTCCACC	ACACAGCGGG
GCTCTGTGTA	GTTTACAAAT	ACCTGTGTGG	AAAGCTGGAC	CGATGTAAAG	GTTCGGGGCT	GTGCTTTTAA	CTGCTGCTGG	AAAGGGGTGG	TGTGTGCCCC
2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
CTTTTCTGTC	CGAAGTTAAT	TCCTTACGGA	CAAACTTTTC	ACATGGAAAC	CATAGAGACG	ACTCCCATTC	AGGTCCCAAG	GGGTGTACAA	CCGGAGGCTG
CAAAAGCAGG	GCTTCAATTA	AGAAATGCGT	GTTTGAAGGG	TGTACTCTTG	GTATCTCTGC	TGAGGGTAAAC	TCCAGGGTGC	GGCAACAATG	GGCTCCGACG
3010	3020	3030	3040	3050	3060	3070	3080	3090	3100
ACACCAACGA	AGTAGACATCA	CTTTTCCGAC	CGACACTAAT	TGCTATTTGA	CCACACACCG	TGACCGCTCC	TGTCCTGGAG	AGTCTACGAC	TGGACGAGCG
TGTGGTGTCT	TCATGCTGAT	GAAGAAGGTC	AGCATTAATTA	AGCATTAACAT	GGTGTGTGGC	AACTGCGAGG	ACAGGGGCTC	TCAGATGCTG	ACTGCTCGGG
3110	3120	3130	3140	3150	3160	3170	3180	3190	3200
TGCGGTGACG	AGTGAACGAC	TCTGTGTAAAG	TGCTATGCTG	GGTGAGAGCG	TTCCGGACCG	GTCAACAACG	CGTGTGTGAT	GACTGGGGCA	CAAGGAAGCT
ACGGCAACTG	TCATCTGTGG	AAGACCATTC	AGGTAGCCAG	CCACTCTCGC	AAGGCGCTGG	CAGTGTGTGA	GCACAACATA	CTGACCCGCT	GTTCCTTGCA
3210	3220	3230	3240	3250	3260	3270	3280	3290	3300
AAACCCATTC	TCCTCCCGCC	ACAAGGATGG	AATGCTTACG	TAAACTCAAG	TGTGATTCTA	TAAACCAATC	GGGCTCTCGT	ACAGGTTECA	CTTGGACTTG
TTTGGGTAAAC	AGGAGGGGGG	TGTTCTATCC	TTACCAATGC	AAATTTAGTC	ACACTAAGAT	ATTCCTTTAG	CCCGAGAGCA	TGTCCAAAGT	GAACCTGACG
3310	3320	3330	3340	3350	3360	3370	3380	3390	3400
CCCCACAAAC	TGTACTGGTA	CTTCTAGACC	TYCCACGACT	CCATGCTACT	CTGGGCTGGG	TCCACGTCTG	GGACGCTCAC	ACCGCCATT	GTATATCTCT
GGGTGTGTTG	ACATGACCAT	GAAGATCTGG	AGGTGCTGTA	GGTACGATGA	GACCCGACCC	AGGTGACAGC	CCTCGGAGTG	TGGCGGTAAA	CATATATAGGA
3410	3420	3430	3440	3450	3460	3470	3480	3490	3500
TGGTGGACAA	CTACGACCTA	CACTGGCTCC	TGCACTCCGG	GCTAGTGAAC	CACGACCGGA	CGTGGGGCGG	ACTCAAAACG	ACATCGCTAC	TTCATGTCTC
ACCAGCTGTG	GATGCTGGAT	GTGACCGAGG	AGCTGAGGCC	CGATCAGTTG	GTGCTGGCCT	GCACCCGCGC	TGAGTTTGGC	TCTAGCGATG	AAGATACAGA
3510	3520	3530	3540	3550	3560	3570	3580	3590	3600
AACTCCATGA	CTTTACACAC	CGCACCGCAA	TTCCCAACCT	TTCTTATATA	TTCCACCCCC	AGAGTACATC	AAAACATAGA	CAAAACGTCC	TGCGCGCGCG
TTGAGGTACT	GAAGTGTGTC	GGCGTGGCTT	AGGGTGGGGA	AAGAAATATAT	AAGGTGGGGG	TTTGTATCT	TTTGTATCT	GTTTTGCAGC	AGCCCGCGCC
3610	3620	3630	3640	3650	3660	3670	3680	3690	3700
ATGACCGCCA	TGACCAAACT	ACCTTCTGTA	CACCTGAGTA	TAACTCTGTC	CGGTACCGGG	GGTACCGGGC	CCCAAGCACT	CTTACACTAC	CCGAGGTGCT
ATGACCGCCA	ACTCGTTTGA	TGGAAGCATT	GTGAGCTCAT	ATTGACAAC	GCGCATCGCC	GGGTGGGCGG	GGGTGGGCTC	GAATGTGATG	GGCTCCAGCA
3710	3720	3730	3740	3750	3760	3770	3780	3790	3800
AACTACCAAC	GGGCGAGGAC	GGGCGTTTGA	GATGATGGAA	CTGATGCTTC	TGGCAGACAG	CTTCCGCGAA	CCCTGTACGT	CGGAGCGCGG	GGCGAAGTCG
TTGATGGTGG	CCCGGTCTCG	CCCGCAAACT	CTACTACTCT	GACCTACGAG	ACCGTGTCTG	GAAAGCGGTT	GGAGACTGCA	GGCTCCGCGG	CCGCTTCAGC
3810	3820	3830	3840	3850	3860	3870	3880	3890	3900
GGCAGCTCGG	TGGCGGGCGG	CCTAACACTG	ACTGAACGCA	AAGGACTCGG	GCGAAGCTTC	GTCACCTCGA	AGGCAAGTA	GGCGGGCGCT	ACTGTCTAAC
CGCTGCAAGC	ACCGCGCGCG	GGATTGTGAC	TGACTTTGCT	TTCTGAGGCC	CGCTTGCAAG	CAGTGCAGCT	TCCCTTTCAT	CCGCCCGCGA	TGACAAGTTG
3910	3920	3930	3940	3950	3960	3970	3980	3990	4000
TGCGGAGAAA	ACCGGTGTTA	CCTAAGAAC	TGGGCTCTTC	AATTACAGCA	AAGAGTCTCT	SACAACTTAG	ACGCGGTCTG	CCAAAGACGG	GACTTCCGAA
DACGGCTCTT	TGGCACAATT	GGATCTTTTG	ACCCGCGAAC	TATATGTCT	TTCTACAGAG	CTGTTGGATC	TGCGGAGACA	GGTTTCTGCC	CTGAAGGCTT

FIGURE 18 (Continued)

4010 GGAGGGAGG CCTCCCTCC	4020 GTTACGCCAA CAATGCCGTT	4030 ATTTTGTATT TAAACATAA	4040 TATTTTGGT ATAAAACCA	4050 CTGAGACAAA SACTCTGTT	4060 CCTAAACCTA GGATTTTGT	4070 GTTCTGTCAC CAAGCAAGTG	4080 AGAACGACAG TCTTGCTGTC	4090 AAATAATCC TTTATTTAGG	4100 CCAAAACGCG GTTTTTGGCG
4110 GGGCCATCC GCGGGTAGG	4120 GGGCCCTGGT CCCGGGACCA	4130 GGCCAGGCC GCGGTCTGG	4140 AGCAACTCC TCGTGAGGG	4150 AGGACACATA TCCGTGTAT	4160 AAAAAGGTCC TTTTTCCAGG	4170 TGCACCATTT ACGTGGTAAA	4180 CCACTGAGAC GGTGACTCTG	4190 CTACAGTCT GATGTTTACA	4200 ATGTACCGGT TACATGGGCA
4210 ATTGGGGAG TAAGCCGTC	4220 AGACCCGACC TCTGGGTGG	4230 TCCATCGTGG AGTAGCACC	4240 TGACGTCTCG ACTGCAGAGC	4250 AAGTACGACC TTCATGCTGC	4260 CCCCACACA GGGGTGGTGT	4270 ACATCTACTA TGTAGATGAT	4280 GGTCAGCATC CCAGTGTAG	4290 GTCTTCGCGA CAGGAGCGCT	4300 CCCGACACAC GGCGGTGGTG
4310 GGATTTTAC CCTAAATATG	4320 AGAAAGTCAT TCTTCACTA	4330 CGTTCGACTA GCAACTGAT	4340 ACGGTCCCGC TCCAGGGGGC	4350 TCCGGGAACC AGGCCCTTGG	4360 ACATTCACAA TGTAAATGTT	4370 ATGTTTCCGC TACAAAGGCG	4380 AATTCGACCC TTAAGCTGGG	4390 TACCACTGTA ATGGGTGCAT	4400 TGCACCCCTA ACGTGGGGAT
4410 TACTCTACGT ATGAGATGA	4420 AGAAGCTGAC TCTTGGACTG	4430 ATAAAATCC TATTTTAGG	4440 AACCAGTACA TTGGCTATGT	4450 AGGGTCCGTA TCCAGGCTAT	4460 TAGGGAGGCG ATCCCTCCGG	4470 CCTAAGTACA GGATTCTATG	4480 ACACGTCTTG TGTGAGAAC	4490 GTGTCGTGT CACCAGACCA	4500 CACATAGGCC GTGTATCCGG
4510 ACGTGAACCC TGCATTTGGG	4520 TTTAAACACT AAATTTGTCA	4530 ACATCGAATC TGTAGCTTAG	4540 TTCCTTTACG AAGGAATGC	4550 CACCTTCTTG GTGGAAAGAC	4560 AACCCTCTCG TTGGAGACGC	4570 GGAAACACTG CCTTGTAACC	4580 AGGCTCTAAA TCCGAGATTT	4590 AGGTAGCTAA TCCATGATTA	4600 CGAGCTATTA GGCTCATTAAT
4610 CTACCTGTAC GATGGCAATG	4620 CCGGGTGCCC GGCCACGGGG	4630 GCCGCCGGAC GGCGGCCCTG	4640 CCGCTTCTAT GGCGAGATTA	4650 AAGACGCTTA TTTCTGGGAT	4660 GTGATTGCA CACTAAGCTC	4670 TATCAACACA ATAGTTTGTG	4680 AGGTCTTACT TCCAGATGTA	4690 CTAGCAGTAT GATCGTCTCA	4700 CCGTATAAAA GGCATATTTT
4710 TGTTTCCGCC ACAAAGCGCG	4720 CCGCTTCCCA GGCGAGGGGT	4730 CGTCTGACG GCCAGACTGC	4740 CCATATTACC GGTATAATGG	4750 AAGGTAGGCC TTCCATCCGG	4760 GGGTCCCGCC CCGAGGGGCG	4770 ATCAATGGGA TAGTTACCTT	4780 GTGTCTAAAC CACAGATTTC	4790 GTAAAGGGTG CATTTCCAC	4800 CGAACTCAA GCTTTGAGTT
4810 GTCTACCCCG CAGATGGGGG	4820 CTAGTACAGA GATCATGTCT	4830 TGGACGCCCG ACCTCGGGGG	4840 GCTACTTCTT CGATGAAGAA	4850 TTGGCAAGAG AACCGTTTCC	4860 CCCCATCCCG GGGGTAGGGG	4870 TCTAGTCGAC AGATCAGCTG	4880 CCTTCTTTTC GGAAGAAAGC	4890 TCCAGAGACT AGGTTCTCTG	4900 GCTCAGCGCT GCAGCTGCGA
4910 CAATGGGCTC CTTACCGCAG	4920 GGCCACCCCG CCGGTGGGCC	4930 GCATTTAGTG CGTAAATCAC	4940 TGGATATGG ACCTATTACC	4950 CCGACGTTGA GGCTGCAACT	4960 CCATCAATTC GGTAGTTAAG	4970 TCTGCACTTC AGAGCTGCAAG	4980 GACGGCAGTA CTGGCGTATC	4990 GGGACTGCTC CCCTGAGCAG	5000 CCCCCGGTGA GGGGGCCACT
5010 AGCAATTCTGT TGTATTAGCA	5020 ACAGGGACTG TGTCTCTGAC	5030 ACGTATCAAA TTGCTATGTT	5040 AGGACTCTGT TCCCTGACCA	5050 TTACGCGGTC AATGCGCCAG	5060 TTCGCGGAGC AAGGCGCTCG	5070 GGCGGCTGCG CCGCCCAAGC	5080 TATGCTCAAG ATAGCATGTC	5090 AGCCTTCTTT TCTCAAGGAA	5100 CGTTTCAAAA GCAAAAGTTT
5110 AGTTGCAAAA TCACGGGTTT	5120 CTCCGGCAGG GAGGCGGTCG	5130 CGGCATCCGT GCCGTAGGCA	5140 ACGAAACTCC TGCCTTTTGA	5150 GCAAACTGGT CGTTTGACCA	5160 TCGTCAAGGT AGCATTTCCA	5170 CCGCGCAGGT GGGCGTCCCA	5180 GTGAGGCGAG CAGCTCGGTC	5190 TGCAGAGAT ACGTGCTCTA	5200 GCCGTAGGCG CGCATCTCG
5210 TAGGTCGTAT ATCCAGCATA	5220 AGAGAGCAAA TCTCTCTGTT	5230 AGCCGCCAAC TCCGCGGTTG	5240 CCGCGCGAAA GGGCGGCTTT	5250 GCGACATGCC CGCTGTACGG	5260 GTCTATCAGC CAGTAGTCGG	5270 ACGAGCAGGT TGTCTGTCCA	5280 CTGCCCGGTC GAGCGGCCAG	5290 CCAGTACAGA GCTCATCTCT	5300 AAGGTGCGCC TTCCACGAGC
5310 CGTCCCAAGG CGAGGGTCTT	5320 GCAGTGCAGT GTCCTGGGTC	5330 CAGACCCAGT GTCCTGGGTC	5340 GCCACTTCCC CGGTGAAGGG	5350 CAGCGCAGGG GCTGCGTCCG	5360 CCGAGCGGCG GGTCTGCGCG	5370 ACCGGTGCCA TGGCCAGGCT	5380 CGCGAATCCG CGGCTTGAAG	5390 GACCGAGGCG CTGCTCTCTG	5400 ACCAAGACTT TGGTCTGAA
5410 CGCGACGGCC GCGGTGCGCG	5420 AGAAGCGGGA TCTTGGGCTT	5430 CGCGACGGCG GCGGCTGCGC	5440 GTCCATCGTA CAGGTAGCAT	5450 AACTGGTACC TGTGATAGTC	5460 ACAGTATCAG TGTGATAGTC	5470 GTGGGAGGCG CACCCTCTCG	5480 GCGACCTGGA CGGCTGCGCC	5490 AACCAGGCTC TTGGCCGAGC	5500 GAGCGGAGAC CTTGCCTCTG
5510 CTCTTCGCGC GAGGAGGCGC	5520 GCGTGTCTCC CGCAGGAGGG	5530 GCTCAGCTCT GCAGTGCAGA	5540 GAAATTTCCC CTTTTAAAGG	5550 GCATCTCGAA CGTAGAGCTT	5560 CCGCGGCTCT GGGCGCGAGA	5570 TTATGCTTAA AATACCGATT	5580 GGCCCTCAT CGCGGAGATA	5590 CCGTAGGCGC GGCATCTCCG	5600 GCGTCCGCGG CCGACGCGCC
5610 GCGTCTGCGA CGCAGACGCT	5620 GAGCGTAAGG CTCGGATTCC	5630 TGCTCGGTCC ACGAGCCAGG	5640 ACTCGAGACC TGAAGCTCTG	5650 GGCAAGCCCG CGGTTCCGGG	5660 AGTTTTTGGT TCAAAAACCA	5670 CCAAAGGGGG GGTTTCCCGC	5680 TACCAAAAGC ATGCTTTTTC	5690 TACCAAAAGC ATGCTTTTTC	5700 ATGCAAGACCA TACCTCTGGT
5710 AAGCTACTCG TTCATGAGC	5720 GCCACAGGTG CGGTGCTCAC	5730 CGAGCCACTG GCTCGGTGAC	5740 CTTTTCCGAC GAAAAGGCTG	5750 AGGACAGAGG TCCGTGCTCC	5760 GCATATGTCT CGTATACAGA	5770 GAACCTCTCG CTTGAAGAGC	5780 GACAGGAGCT CTGTCTCTCA	5790 CGCCACAAGG GCGGCTCTCC	5800 CGCCAGAGGG GCGGCTCTCC
5810 AGCATATCTT TGTATAGAA	5820 TGAGCTGCTG ACTCGAGCCA	5830 GAGACTCTGC CTCTGAGAGC	5840 TTCGAGCGCG AAGGCTGCGG	5850 AGGTCGCGTC TCCAGGCCAG	5860 GTGCTTCTCT CACGAAGGAG	5870 CGATTACCCC GCTAAGTGGG	5880 TCCCATCTCG AGGGGTAGCG	5890 CAGCAACAGG GTGCTGTGCC	5900 TGATCCCGCA ACTAGGGGGT
5910 GGTGAGGAG CCACTGCGCT	5920 GTCCACACT CAGGGTGTGA	5930 TCTGTGTACA AGACACATGT	5940 CGCGGAGAGG CGCCCTCTTC	5950 CGGTAGTCTC GGCATCAAGG	5960 TTCCTAATAC AAGGTGATTC	5970 CAATATATCA GTTTATAGTT	5980 CATCCGCTCC GTAGGCCACG	5990 ACTGCGCCAC TGACGGGGTG	6000 AAGGACTTCC TTCTGAGG

FIGURE 18 (Continued)

6010 CCCCGATATT GGGCGTATAA	6020 TTCCCCCACC AAGGGGGTGG	6030 CCCGCGCAAG GGGCGGCTTC	6040 CAGGAGTAG GTCCCTACTC	6050 AGAAGGCGTA TCTTCCGCAT	6060 GCGACAGACG CGCTGTCTGC	6070 CTCCCGGTCG GAGGGCCAGC	6080 ACAAECCEAC TGTGGGGTGG	6090 TCATGAGGGA AGTACTCCCT	6100 GAGTTTTCGC CTCAAAAGCG
6110 CCGTACTGAA GGCATGACTT	6120 GACGCGATTC CTGCGGTAGG	6130 TAACAGTCAA ATTGTCACTT	6140 AGGTTTTCGC TCCAAAGCG	6150 TCCCTCTAAA AGGAGATTCT	6160 CTATAACTGG GATATTCCAC	6170 ACCGGCGCCC TGGCCCGGCG	6180 ACTACGGAAG TGATGCTTTT	6190 CTCCACCGGG GAGGGTGCGC	6200 GCGAGGTAGA GGGTGCTATC
6210 CCAGTCTTTT GGTCAGAAAA	6220 CTGTTAGAAA GACAATCTTT	6230 AACAACAGTT TTGTTGTCAA	6240 CGAACACCGT GCTTGGTGGC	6250 TTTGGTGGGC AAACGACCGC	6260 ATCTCCCGCA TAGAGGGCGT	6270 ACCTGTGCTT TGGACAGCAA	6280 GAACCGCTAC CTTGGCGATG	6290 CTCGGCTCCC GAGCGCAGGG	6300 AAACCAAAAA TTTGTGTTTT
6310 CAGGCGTAGC GTGCGATCGC	6320 CGGCGGAGGA GCGCGCTCTT	6330 ACGCGGCTTA TGGCGCGGAT	6340 CAAAATGAGC GTTTAGCTGC	6350 TGCAATAAGC ACGTATTCTC	6360 CGGCTGTGCT GCGCAACGCA	6370 CGGCTGTGCT CGGCAATTCG	6380 CGTTTCTGCG GGAAAGACGG	6390 ACACGCGGAG TGGTGGCTCT	6400 CAGCGGCTGG GTGCGGCAAC
6410 TCCAGCTGCG AGGTGACGCG	6420 CGGTTGGCGC GCCAACCGCG	6430 CAACACGTCC GTTGTGACGG	6440 CACTGTTCCA GTGACAAAGT	6450 GTTGCGACCA CAACGCTGGT	6460 CCGATGGAGA GGCTACCTCT	6470 GGCGCATCCG CGGCTATAGC	6480 CGAGCAACCA GCTCGTTGGT	6490 GGTGTCTCTC CCAGCAGAGG	6500 GCCGCGGGGA CGGCGGCTCT
6510 ACGCGCTTCT TGGCGGCAACA	6520 CTTACGCGCA GAATGGCGGT	6530 TCACCCAGAT AGTGGGCTCTA	6540 CGACGAGAGG GCTGCTCTCT	6550 CAGCGCGCCC GTCCGCGGGG	6560 AGACCGAGGT TCTGCTCCCA	6570 GCCATTCTCT CGGTAAAGAC	6580 GGGCGCTGCG CGGCGCGCGT	6590 TCCGCGCGCA AGGCGCGCGT	6600 GCTTCTCTAG CGAAGTAGTC
6610 ATAGAACGTA TATCTTGCAAT	6620 GGAGCGGTTCA CTTCTCAAGT	6630 GATCGCGGAC CTAGCGGCTG	6640 GACGCTACCG CTGCGGCTCT	6650 GCGCGCGGTT GCGCGGCTCT	6660 CGCGCGGAGG GCTGCGGCTCT	6670 CATACCCCAAC GTATGGGTTG	6680 TCACCCCTCT AGTGGGCGAC	6690 GGGTACCGTA CCCATGGCAT	6700 CCCCACCCAC GGGCTGGGTT
6710 TCGCGGCTCT AGCGCGGAGG	6720 GCAATGACGG GCTACATGCT	6730 CGTTTACAGC GCAAAATGCTG	6740 ATTTGCACTG TAAAGTAGTA	6750 CCCCGAGAGA GGGCTCTCTCT	6760 CTCATTAAGT GAGTATTCCA	6770 TCTATACATC AGATATGTAG	6780 CGATCTCTAG GGTAGCTCTCT	6790 AGGTGCGCGC TCCACCGCGG	6800 TACGACGCGG ATGCTGGCGC
6810 CGTGCATTAG GCACGTATTC	6820 CATATCAAGC GATATGTTCT	6830 ACGCTCCCTC TGGAGGGGAG	6840 GCTCCTCCAG CGAGGAGGTC	6850 CCCTGGCTCC GGGACCGAGG	6860 AACGATGCCC TTGCTACGGG	6870 GCGCGACGAG CGGCTGCTCT	6880 ACGAGCTTTC TGTCTGGAGG	6890 TGATAGACGG ACTATCTGCC	6900 ACTTCTACCG TGAAATGGCG
6910 TACACTCAAC GATGAGGTTG	6920 CTACTATATC GATGATATGG	6930 AACCTGCGAC TTGGACGCTG	6940 CTTCTGCAAC GAAGACGTTG	6950 TTGACCGGCA AAGCTGGCGT	6960 GACACTCTGG CTGTGAGACC	6970 ATGGCGCAGT TACCGCGTCA	6980 CGGTGTCTTC CGCACGAAGG	6990 TCCGCTCTCT AGGCGTAGGA	7000 CAGCGGCTCG GTGCGGCAAC
7010 AACAACTGGT TTGTTGACCA	7020 CGAGCGGCGA GCTGCGGCTG	7030 CTGGACGTGC GACTCTGACG	7040 AGATCCCGCG TCTAGGGCGC	7050 TCATCAGGTC AGTAGTCCAG	7060 CCAAAGGAAC GGTTTCTCTG	7070 TACTACAGTA ATGATGCTAT	7080 TGAATAGGAC ACTTATCTCT	7090 AGGGAAAAAA TCCCTTTTTT	7100 AAGGTGTCGA TTCCACAGCT
7110 GCGGCACTGC GCGGGTTGAG	7120 CTGTTTGAAG GACAACTCTT	7130 AGCGCGAGAA TCCGCGTCTT	7140 AGGTCATGAG TCCAGTACTC	7150 AACCTAGCCT TTGGATCGGA	7160 TTGGGCGAGC AACCCGTGCG	7170 GGAGGCTTGC CCTCCGAAAC	7180 CATTCGCGGA GTAAAGGCTT	7190 TCGTACATCT AGCATGTAGA	7200 TGACCAACTG ACTGGTTGAC
7210 CGGAGCAATC GGCCTGCTAG	7220 CGGCTGCTAG GCGCAGCATC	7230 GGAAGAGATG CCTTTTCTAC	7240 CCATCTCGCG GGGTAGCGCG	7250 ATAGCGACGC TATGCTTGCG	7260 GCGCGAAGGC CGGCTTCTCG	7270 CTGCTCTCAC GAGCGAGGTT	7280 ACCCACTGCG TGGGTGAGCG	7290 GTTTCCACAG CAAGGTTGTC	7300 GGATTGGTAC CCTAACCATG
7310 TGAAACTCCA ACTTTGAGGT	7320 TGACCAATAA ACTGGTATTT	7330 CTTCAGTCAC GAAGTCAGTG	7340 AGCAGCGTAG TCGTGCGATC	7350 GCGGAGCAGG CGCCCTGCTC	7360 GGTCTCGTTT CCAGAGCAAA	7370 TTCAAGGACG AAGTCCGTGC	7380 CGAAAAACCT GCTTTTGGGA	7390 TGCGCCCAAA ACGCGGGTTT	7400 CCGTCCCGCT GGCAGGGCGA
7410 TCCACTGTAG AGGTGACATC	7420 CAACTTTTCA GTTGAAAAAT	7430 TAGAAGAGGC ATCTTTTCCG	7440 GCGCTCGGTA CGCGAGGCAAT	7450 TTTCAACGCA AAAGTTGCGT	7460 CACTACGCTT GTGATCGGGA	7470 TCCGAGGCGC AGGGTCCCGG	7480 GTGGAGGCTT CACCTCGGAA	7490 GCCAACAATT CGGTTGTAAA	7500 AATGAGCGGC TTACCTGGGC
7510 GCGCTGCTGC GCGGAGCAGG	7520 TAGAGCAGCT ATCTGCTCGA	7530 TCGGCAACTA AGCGTTGAT	7540 CAACACCGGG GTTGTGGCGC	7550 TGCTACATTT ACGATGTAAA	7560 CAAGGTTCTT GTTCCAAGAA	7570 GCGCGCCACG CGCGGGGGTG	7580 GGGAACCTAC CCCTTGATGG	7590 TCCCGTTAAA AGGGCAATTT	7600 AAATCAAGG TTTAACTTCC
7610 AGCATCCACT TCGTAGGTGA	7620 CGAGGAGTCC GCTCTCAGG	7630 CCTCGACTCG GGAGCTGAGC	7640 GGCACAAGAC CGGTGTTCTG	7650 TGTCGCGGCT ACAGGGCCCA	7660 CAGAGGTTCT GTCTGCAAGA	7670 ACTCCCAACC TGAGGGTTGG	7680 TTGCTGCTTT AAGGAGCGAA	7690 ACTCGAGGCT TGAGCTCCAC	7700 TCCAGTGGCC AGGTACCGGG
7710 GGTAATCGTA CCATTAGCAT	7720 AACGCTTACC TTGCGAGTGG	7730 AGCGGTTTCC TGGCGAAGGG	7740 AGGATTTGAC TCTTAAACTG	7750 CGCTGGATAC GCGACCTATG	7760 CGGTAAAAAA GCCATTTTTT	7770 GACCCCACTA CTGGGGTGAT	7780 CGTCACTTTC GCAGTAGAAG	7790 CATTCGCCCA GTAAGCGGGT	7800 GAACAAGGGT CTGTCTCCCA
7810 CGCGAGGCTA CGGGTCCCAT	7820 GGTTCAGGTT CCAAAGTCCA	7830 GCCGATCCAG CGGCTAGGTC	7840 AGCGCGCGCG TCCGCGCGCG	7850 CAGTGGTCTC GTCAACAGAG	7860 CGAGTAGAGG GCTCATCTCC	7870 CGGCTTGAAG GCGCAACTTC	7880 TATGCTGCTG ATAACAGAGA	7890 ACTTCCGCTG TGAAGGGCAC	7900 CTCGACGAAG GAGCTGCTTC
7910 GGTTTCCGGG H CCAAGGCGCC	7920 GGTAGGTTCA CCATCCAAAT	7930 TATCCAGAGA ATAGGTTCTT	7940 TGTAAGCATCC ACATCGTAGG	7950 ACTGTTTCTC TGACAAGAG	7960 TGCAGGCCAC ACGCTCGGTT	7970 GCTCTACGCG CGAGGATGCG	7980 TCGGCTAGCC AGCGGCTCGG	7990 CTTCTTGACC GAAGAACTGG	8000 TAGAGGGCGG ATCTCCCGCC

FIGURE 18 (Continued)

8010 TGGTAACT ACCAATTGGA	8020 CCTACCGAC GGAGTGGCTG	8030 AACTACACCA TTGATGTGGT	8040 CTTTTCATCTT GAAAGTAGAA	8050 CAGGAGACGT GTCCCTGCGA	8060 GCCCGGCTTG CGGGCCGAA	8070 TGAGCAGGAC ACTGCTGCTG	8080 CGAAAACATT GCTTTTGTAA	8090 TTTGCACGG AAACGTGGCG	8100 TCATGACGCT AGTACTGGCA
8110 CGCCACGTGC GCGGTGCAAG	8120 CCGACATGTA GGTGTACAT	8130 GGACGTGCTC CTGCGACGAG	8140 CAACTGGACT GTTGACCTGA	8150 GCTGGCGGT CGACCGCGEA	8160 GTTCTTCTGT CAAGGAAGCA	8170 CTACCCCTTA GAGTGGGAAT	8180 AACTCGGGGA TTGAGCCCTT	8190 GGGACCCGCC CGCTGGCGCG	8200 CAAAACGACC GTTTGGCTGG
8210 ACCAAGAGAT TGGTCTCTTA	8220 GAAGCGGAC CTTCCGCTGC	8230 AACAGGAAC TTGTCTTGA	8240 GGCAGACCGA CGCTCTGGCT	8250 CGAGCTCCCG GCTCGAGGGG	8260 TCAATACCA AGTATGGTGG	8270 CTAGCCTGGT GATCGGACCA	8280 GCTCGGGGCG CCAGCGCGCG	8290 GCTCGGGTTT CGAGCCCAAA	8300 CAGGTCTACA GTTCCAGATG
8310 GCGCGCGCGC CGCGCGCGCG	8320 GCCAGCTCTG CGGTGCGAGC	8330 AACTACTGTT TTGATGACAA	8340 GTAGCGGCTC CATCGCGCAG	8350 TACCGCTGAC ATGGGAGCTG	8360 AGGTACAGAA TCCATGGTCT	8370 CCTCGAGGGC GGAGCTCCCG	8380 GCCGTGTGTC CGCGGACAGG	8390 AGTCCGCCCT TCAGCGGGGA	8400 CGAGGACGTC GCTCTGCGAG
8410 CAAAATGGAGC GTTTACCTCG	8420 GTATCGGCC CAATCGCGGG	8430 AGTCCCGCGC TCAGCGCGCG	8440 CCGATCCAGG GGCTAGGTCC	8450 TCCACTATGG AGGTGATACC	8460 ACTAAGGGTC TGATTTCAGG	8470 CCCGACCAAC GGGCTGGTTG	8480 CACCGCGCGA GTGGCGCGCT	8490 GCTACTGAAC CGATGACTTG	8500 GTTCTCGCGC CAAGAGCGCG
8510 GTAGGGCGCG CATCCCGCGC	8520 CGCGCTGATG GCGCGACTAC	8530 CCATGCGCGC GGTACCGCGC	8540 CGCGCGCGCG GGCGCGCGCT	8550 CGCGCGCGCG GGGTCTCTTG	8560 CCACAGGAAC GGGTCTCTTG	8570 CTACTACGTA GATGATGCAT	8580 GATTTTCGCC CTAAAGCGGG	8590 ACTCGCGCGC TGACCGCGCG	8600 CGCGCGCGCG GGGCGCGCGC
8610 TCCATCCCCC AGGTAGGGGG	8620 CCGAGCGCTG GGCTCGGGAG	8630 GGCGCGCGCT CGCGCGGGAG	8640 TCCCGCGCTC AGGGCGGAGG	8650 CGGTGAGGCC GGGCGCGCGC	8660 GGCGCGCGCG CGCGCGCGCG	8670 CCCGTCTCTG GGGCGAGGAG	8680 ACCAGGAGCG TGGTCTGCGC	8690 CGCGCTCCAA CGCGGAGGTT	8700 CGACCGCTTG GCTGGCGAAC
8710 CGCTGCTGCG CGCGAGCGCG	8720 CCCGCAACTA GGCGGTTGAT	8730 GAGGACTTAG CTCTGGAATC	8740 ACCGCGGAGA TGGCGCTCTT	8750 CGCACTTCTG GGCTGAAGAC	8760 CTCGCGCGCG GACGGCGCGC	8770 CACTACGAAC GTGAGCTTGA	8780 TGGACTTTCT ACCTGAAGAA	8790 CTCAAGCTGT GAGTTCGACA	8800 CTTACTTAA GAATCAATTT
8810 GCCACAGCAA CGGTGCTGTT	8820 CTGCCCGCGC GACGGCGGCC	8830 ACCGCGTTTT TGGCGCAAAA	8840 AGAGGACGTC TCTCTGCGAC	8850 CAGAGGACTC GTCTCTGAGG	8860 AACAGACACTA TTGCTTTGAT	8870 TCCGCTCAAG AGGCGATTTC	8880 CGGCTACTTG GGCCATGAAC	8890 ACGAGCTAGA TGTCTGCTCT	8900 GAAGGAGGAC GTTCTGCTCT
8910 CTCTAGAGGG GAGATCTCGG	8920 CGAGCGCGAG CGTCCGGCTC	8930 CGAGGTGCTG GCTCCACGGT	8940 CGCGCGCTCC GGCGCGGAGG	8950 AGCAACCTCT TGTGTTGAGA	8960 ACGCGCGGTA TGGCGGCGAT	8970 CTGAGCGCTC GAGCTGCGAG	8980 TTCGCGAACT AGGCGTTTGA	8990 CCCGAGGGAG GCGCTCCCTC	9000 CAAGGTGCTC GTTCTGAGCG
9010 CCCGACATCT CGCTGTGAGA	9020 GGTCCGGGGG CCACGCGCCC	9030 AAGCCCTAGC TTCCGGCATCG	9040 GCCCGCGCGT CGCGCGCGCA	9050 ACTGCTGGAC TGACCACTCTG	9060 CGCTCTTAAC AGCTCCACGT	9070 TCGAGGTGCA CGCGAGATTG	9080 CGCGCGGCTT GCCGGCGGAA	9090 CTGCCGCGATC GACGGCGTAG	9100 AAAGCTCGCC TTTCTGAGGC
9110 CGACTTTTCT GCTGAAGAGG	9120 CATCAACTCC GTAGTTGAGG	9130 CACCAACGCC GTGGTGGCGG	9140 ACACAAAGCG TGTGTTCTGC	9150 GTGCTTCTTC CACGAAGAAG	9160 ATGATTTGGG TACATAACCC	9170 TCCGAGCGTT AGGCTGCGAA	9180 GCACCTAAGC CGTGGATTGG	9190 AACTATAGGG TTGATATCCC	9200 GTTTCCGGAG CCAAAGGCTC
9210 TTCCGCGAGG AAGGCGCTCC	9220 TACCGGAGCA ATGCGCTCGT	9230 TCTTCAGGTG AGAAGTCCAC	9240 CCGCTTCAAC GGCGGAAGTT	9250 TTTTTGACCC AAAAACTGGG	9260 TCAACGCGCG AGTTGCGCGC	9270 GCTGTGCCAA CGACAGCGTT	9280 TTGAGGAGCA AACTCTCTCT	9290 GGTCTTCTCG CCAGAAAGCG	9300 CTACTCGAGC GATGAGCTCG
9310 CGCTGTCAACA CGCAGAGTGT	9320 CGCGGTGGAG CGCGCACTCT	9330 CGCGAGTTTC CGGCTCAAGG	9340 CGATGTCCCG GCTCTTCTTC	9350 GGAGAGAAGG CTCAATCTCC	9360 AAGTTAGAGG TTCAATCTCC	9370 AGAAGGTATT TCTTCCATAA	9380 CCCGAGGGGG GGGCTCCCGC	9390 AAGAAGAGCA TCTTCTCTCT	9400 AGAAAGCCCG TCTTCTGGCG
9410 CGCCACCGCC CGGCTGGGGG	9420 TCCCGCTGCT AGGGGGGACA	9430 GGCGCGCGTG CGCGCGCGAC	9440 CTGCGCGGTG GACGGCGCAC	9450 GCCCTCCCGC CGGGAGGCGG	9460 AGCTGTTTTC TCGACAAAGC	9470 CGAGCTAGTA GCTGATCAT	9480 GAGGGCGGCC CTCCCGCGCG	9490 GCTGCCCGGT CGAGCGCGCA	9500 ACGAGAGCCA TGGTCTCGGT
9510 CTGCCCGCGC GACGGCGCGG	9520 GGCAAGAGCG CGTTTCTGCG	9530 CCCCCGCGTC GGGGCGCGAG	9540 AACTTCTCTG TTGGAAGAGC	9550 CGCGGGGAGT CCGCGCGTCA	9560 ACAGGGGCCAA TGTCCCGGTT	9570 TACCCAAACG ATCGGTTGCG	9580 CCCCCGGAGC GGGGGGCTCG	9590 GCACGCGGTC GTCGCGCGAG	9600 CCTATGCCCG GGATAGCGCG
9610 GATGCTAGCG CTAACGATGC	9620 TAGAGTGTGT ATCTCAACAA	9630 AACACACAT TTGTTGTGTA	9640 CCATGAGGCG GGTACTCCGC	9650 GTGGCTCCCT CACCAGGGGA	9660 GGACTCGCTC CCTGAGCGAG	9670 ACCGCTACTC TCCCGCATGA	9680 GGCCTAACCT CGGATCGGGA	9690 TTTGGAGAGC AAACCTCTCG	9700 TCTTCCCGCA AGTAAAGGCT
9710 GATTGGCTCAG CTAACGATGC	9720 TGTACGGGTT ACAGTGGCAA	9730 CCATCCGACT GCTAGGCTGA	9740 CGTGGCGAGC GCACCGTGGC	9750 CCCGCGCGTC GGGTGGCGAG	9760 CCCACCGCCA GGGTGGCGAG	9770 GCCCAACCAA CGGGGTTGTT	9780 AGACCGGCTC TCGGCGGAGG	9790 CAGCAGCACT GTGCTGTGTA	9800 ACTACATTA TGATGTAATT
9810 TTTCTCCCGC AAGATGAGCG	9820 CAGAACTCTG GTTCTGAGAC	9830 CGCGCTACCA GGCGGATGTT	9840 GCTGTCTTTC CGACAGAAGC	9850 TGTACAGGTA ACCATGTCTT	9860 ACCCAGGCGC TGGTCCCGGC	9870 GACGACTTAC CTGCTGAATG	9880 GCGTCCCGCA CGCAGCGGCT	9890 GCCGCTACGG CGGCAATGCC	9900 GGTCCGAAGC CGAGGCTTTC
9910 J AAAACTGTAG TTTTGACATC	9920 CGCGGTCGAG GGCGGAGTGC	9930 AAACATCATC TTTGTAGTAG	9940 AGAACGTACT TCTTGCATGA	9950 CGGAAGATG GCCTTTCTAC	9960 GCCGTGAAGA CGGCACTTCT	9970 AGAAAGGAGAA TCTTCTCTCT	9980 GGAGAACAGG CCTCTGTGTC	9990 ACGTAGAGAA TGTATCTCTT	10000 CGTAGATGCG GCATCTATCG

FIGURE 18 (Continued)

FIGURE 19. Nucleotide sequence of a region between coordinates 49.0 and 51.8 on the Ad2 genome. This sequence was determined by Akusjärvi and Persson (1981a). The positions of strategic signals were determined by Akusjärvi and Pettersson (1979a) and Akusjärvi and Persson (1981a).

10010 10020 10030 10040 10050 10060 10070 10080 10090 10100
GATGCCGCG CCCTCTCAAA CCCTGAGTT CCCTGAGTT AGGAGGCTAC GCACACTGGG GCTTCGGGGA GTAGCGGCT TCGTCCGGT CCAAGCGCTG
CTAGCGGCG GCGCGAGTT GCGCTGAGT GCGCGCTCT TCTTCCATG CGTGTGACCC GAGAGCCCTT CATCGGCTA AGCAGGCGCA GGTGCGGCG

10110 10120 10130 10140 10150 10160 10170 10180 10190 10200
TTCGGGAGC CATTATATC GGACGAGCT GACGACTCC CATCTGACT TCACTAGGTA CAGGTGTTT GCACATATC GCGGCGACA CTACCACTT
ACCGCGCTG GCTAATATG CCGCTGAC CCGCTGAGG GTAGACTGA AGTCACTCAT GTCCACAA GCGGTGTATG CCGCGCTGTT SATGTGTATA

10210 10220 10230 10240 10250 10260 10270 10280 10290 10300
CACGTCAAC GGTATTGCT GGTCAATTG GGTCAATTG GCGCGAGCT CTGAGGCEAC ATGGAATCTG GCGTATTCT GGAACCTAGT TTCGTGCTA
GTGCAATTG CCAATACGA CCAATACGA CCAATACGA CCGCTGCGA GAGCTGCGT TACCTGAGAC GCGAGTAGC CTTTGAATG AAGCGGTAGT

10310 10320 10330 10340 10350 10360 10370 10380 10390 10400
GCAAGCTTA GGTGCTGCT ATGACTATG GTGCTGTTT CACGCGGCG CCACGCGGCA TCTCCCGGT GCGATCCCA CCGCGCGGAG GCCCGCGCT
GCTTCAAGT CCGCAGGAG TACTGATAT CCAACAAAA GTGCGGCGG GCGTGGCGT AGAGGCGCA GCGTASGCT GCGCGCGCT CCGCGCGCT

10410 10420 10430 10440 10450 10460 10470 10480 10490 10500
CAGAAAGTT TATTCGCTA CTATAGGAT CTATAGGAT CTGATGCTT ACTACGCGC CCGCGCGGCT CCGCGCGGCT CTTTCAAGC CTGCGCGAG
GCTTCAAGT ATAGGCTAT GATATGCTA GATATGCTA GATATGCTA GATATGCTA GATATGCTA GATATGCTA GATATGCTA GATATGCTA

10510 10520 10530 10540 10550 10560 10570 10580 10590 10600
GCTTCAAGT GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
CAGATGTTG CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

10610 10620 10630 10640 10650 10660 10670 10680 10690 10700
ACATTCGCT GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
TGTAAAGG GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

10710 10720 10730 10740 10750 10760 10770 10780 10790 10800
GTAGCGGCT CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
CATCGGCTA CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

10810 10820 10830 10840 10850 10860 10870 10880 10890 10900
CGAAAGGCT GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
GCTTCAAGT CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

10910 10920 10930 10940 10950 10960 10970 10980 10990 11000
CAATTCGCT GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
GTTGCTGCT CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

11010 11020 11030 11040 11050 11060 11070 11080 11090 11100
GCTTCAAGT GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
CGAAAGGCT CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

11110 11120 11130 11140 11150 11160 11170 11180 11190 11200
GCTTCAAGT GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
CAGATGCTA CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

11210 11220 11230 11240 11250 11260 11270 11280 11290 11300
GCTTCAAGT GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
CGCGCGGCT CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

11310 11320 11330 11340 11350 11360 11370 11380 11390 11400
GCTTCAAGT GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
CGCGCGGCT CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

11410 11420 11430 11440 11450 11460 11470 11480 11490 11500
GCTTCAAGT GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
CGCGCGGCT CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

11510 11520 11530 11540 11550 11560 11570 11580 11590 11600
GCTTCAAGT GGTGCTGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG
ACCTGCTAC CCGCGCGCT TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG TTTCAAGG

FIGURE 18 (Continued)

10 20 30 40 50 60 70 80 90 100
13 GCGCTTACG TAGCGAGCG AACGTGCGG TCTGTGTGAC TAATTTTGT TCAACATGCA CTTTATTATG TTTATTATG AGACCTCAGA GTGCGAGGCA
15 CCGGAAATGC ATCTGCTGCG TTGCGAGGCG AGAGACACTG ATTAATAACA AGTTGCTGAT GGAATAATCA ATATATAAAG TCTGAGTGT CACCTGCGCT

110 120 130 140 150 160 170 180 190 200
ACCAGGACAT TGATAAACA TCTTACTCTT TGTAGTTGAA ACCGAGAGAC CCGGCGCGCTG TGCCGAGCGC GGGCAAGTAC CTTTGAACG TTCTATAGCG
TGCTCTGTA ACTATTGTT AGAATGGAAG ACATCAACTT TGGCTCTCTG GCGCCGCGAC ACGGCTCGCG CCGCTTCATG GGAACATGCG TTCTATAGCG

210 220 230 240 250 260 270 280 290 300
GCGGTGCTTA TACTGCGCAC CCGCGAGGTC GACCACTGCG CGTAATTTT AAAGCCAAAG TGTAAATCT TGTATCCGTC GTTCCGAGCC
CATAGGAGT ATGAGCGGTG GCGCCTTCAG CTGCGGCTCG GCAATTAATA TTTGCTTCC ACCATAAGA ACTATGCGAG CTAAGGCTGCG

310 320 330 340 350 360 370 380 390 400
TTGTGCTGCT GTCCGCTCTA CCACTCCCTG TCAACTTTCG TCGTTTAAAG TCGTTTAAAG GGTGTTTTC GGTGTTTTC CACCACTAC CCCCACCAAC
ACAGCAGCA CAGGCGAGAT GCTGAGGAGC AGTTTGAAG AGCAAAATTT GGTGTTTTC CCAACAAAG GTGTGATG GCTTGTGATG TGGCATTAGC

410 420 430 440 450 460 470 480 490 500
TGAGCCGCTT GGTGCTGCT CCAAGGAGTG GTTTTATCTT AATTGCTATT CGAAGTAGGG GCGGGAGGCG ATCTCTCTCG AGGTGCGCG CACCTCTGTC ACAGAGGCTT
ACCTGCGCA GGTGCTGCT CCAAGGAGTG GTTTTATCTT AATTGCTATT CGAAGTAGGG GCGGGAGGCG ATCTCTCTCG AGGTGCGCG CACCTCTGTC TGTCTCCAGA

510 520 530 540 550 560 570 580 590 600
CGCGCGGCTG CTTTTCGCG GCGCGCGGCT GCGCGCGGCT GCGCGCGGCT GCGCGCGGCT GCGCGCGGCT GCGCGCGGCT GCGCGCGGCT GCGCGCGGCT

610 620 630 640 650 660 670 680 690 700
GCGGTGCTGCG CAGGCTGAGG CCGGTACCGA TGGCTCTCAC ACCCGGCTGT GTGTGACAT TCGGACCTTG ACAGAGGCGG CCGCTGACCC GTCGCTTTTG
CGCATCAACC GGTCCATGCG GCGGTACCGA TGGCTCTCAC ACCCGGCTGT GTGTGACAT TCGGACCTTG ACAGAGGCGG CCGCTGACCC GTCGCTTTTG

710 720 730 740 750 760 770 780 790 800
GACACGACCG TCCCGGACG CCGCGGACG ATTTGCGGCG GCGCGGCGG GCGCGGCGG GCGCGGCGG GCGCGGCGG GCGCGGCGG GCGCGGCGG

810 820 830 840 850 860 870 880 890 900
ACCGTGTGAC GTTTCGCTG ACTTGTGCTA GCAACGAGAT CCGCGGCTGT GCGCGGCTGT GCGCGGCTGT GCGCGGCTGT GCGCGGCTGT GCGCGGCTGT

910 920 930 940 950 960 970 980 990 1000
ACATACGCTG GTATACGCG GGTCTCTCTG ACAGTACGCG GCGCGGCTGT GCGCGGCTGT GCGCGGCTGT GCGCGGCTGT GCGCGGCTGT GCGCGGCTGT

1010 1020 1030 1040
TGATAGGCGG GTCTCTGCG AGCTCTATG ACTGCGGCGG

10 20 30 40 50 60 70 80 90 100
 15 3' TAGGATACG TCCTCGGGTC GGAAGAATA CAAAACAAAC TTCAGAACT GCACACGCA CACGTGGTC CCGTGGCC GCAGTACATC TGGCACTATG
 16 5' GATCCCATCG ACAGGCCAC CTTCTTTAT GTTTTGTTC AATCTTTGA CCGTGGCCG GTCCACACG CCGTGGCCG CCGTGGCCG CCGTGGCCG
 110 120 130 140 150 160 170 180 190 200
 ACCGCTGGCG GAAGAGCCCG CCGTGGCGT GTTGATTTT CTTCGTTCT TGTAGTTGT GTCCACGGCG GTACCCGAGG TCACCTCTCC TTGACTTTTC
 TCGCCAGCGC CTTCTGGCC GGAACGCCA GTTGTATTT CAACATAAAA GAAGCAAGCA ACATCAACAA CAGCTGCCCG CATGGCTCCC AGTACGACG
 210 220 230 240 250 260 270 280 290 300
 GTAACAGTTT CTAGAACCA CACCCGGAT AAAAACCCT TCGATACGT TCGCAAAAG AGAGGTGTGT TCAGCGGAG TCAGCGGAG CCGGTATCAG
 CATTTGCAAA GATCTTGGT GTGGGCCATA TTTTGTGGC ACCTATGAC AGCGCTTTC AGGCTTTGTT TCTCCACACA AGCTCGCTCT CCGCATATGC
 310 320 330 340 350 360 370 380 390 400
 TTATCGCCGC CAGCGCTCTG ACCCGCCAT GTGACCTACC GGAACCGAG CTTTGGCTG AGTTTTTGT CCGATGGAGA ACTCGCGAAA CCGAAGAC
 AATACGGCG GTCCGGACG TGGGGCGTA CACTGGATG GAAACCGAG GAAACCGAG GAAACCGAG GAAACCGAG GAAACCGAG GAAACCGAG
 410 420 430 440 450 460 470 480 490 500
 TGTTCCTGTA GTTCGTGCA ATGCTCAAC TCATGCTAG TGAGGACGG GCATCGCGT AACGAAGAG GGGGCTGGC ACATATTGCG ACCTTTTCAG
 ACGAAGCGT CAAGCAGGT TACCATTTG AGTAGAGTG ACTCTGCGC CTATGCGCA TTGCTTCTC TCGGCTGGC CCCCAGCGC ACATATTGCG TGTATACGC
 510 520 530 540 550 560 570 580 590 600
 GTGGTTTCG CACGTCCCG GTTGAGCGG CCGGACACC GATAAGCA CCGTAAAGA CCGTAAAGA CCGTAAAGA CCGTAAAGA CCGTAAAGA
 CACCCAAAGC GTGCGGGCG CCAACTCGG CCGCTGTGA CTATTCTGT GCATGTTCT CCGTAAAGA CCGTAAAGA CCGTAAAGA CCGTAAAGA
 610 620 630 640 650 660 670 680 690 700
 TTGGGTTGT ACTTGAATA ATGGCCCAT GGGTTGAGT ACGAATTGT TCGTTAAGC AGGGGTCAT AGGGGTCAT AGGGGTCAT AGGGGTCAT
 AATCCCAACA TGAACCTTAT TACCGGGTA CCGCATCCA TCGTTAAGC TCGTTAAGC TCGTTAAGC TCGTTAAGC TCGTTAAGC TCGTTAAGC
 710 720 730 740 750 760 770 780 790 800
 AGACCTTCC GGTGAGCGG ATGAAGCGT CCGTGTACG CGTCTAATC TCGCGGTGA GAAACAGAG GAACCTTTTG TACATTTTGA TTACATGTC
 TCGTGGAGG CCACTCGCC TACTTCGCA TCGTGTACG CGTCTAATC TCGCGGTGA GAAACAGAG GAACCTTTTG TACATTTTGA TTACATGTC
 810 820 830 840 850 860 870 880 890 900
 CTCTGTGAAA GTTATTTCG TTTCAAAAA TAAACATGT ACACCGACT AATAAAGAG CCGTAAAGA CCGTAAAGA CCGTAAAGA CCGTAAAGA
 GAGACATTT CAAATAAGGC AATGTTTTT TAAACATGT ACACCGACT AATAAAGAG CCGTAAAGA CCGTAAAGA CCGTAAAGA CCGTAAAGA
 910 920 930 940 950 960 970 980 990 1000
 A AGACGCGCG TAGCGATAC CCGTACCGT CCGTGTGCA CCGTGTGCA CCGTGTGCA CCGTGTGCA CCGTGTGCA CCGTGTGCA
 ATCTGCGCG ATCGCTATG GCGACTGGA CCGTGTGCA CCGTGTGCA CCGTGTGCA CCGTGTGCA CCGTGTGCA CCGTGTGCA CCGTGTGCA
 1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
 CTTCAAAAGT GAGGTGTCC ACCTGCTGA GTGGTGGCG AATCTGTCA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 GAGTTTTC CTTCCAGGC TCGCAACAT GTGGTGGCG AATCTGTCA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
 CTTCAAAAGT GAGGTGTCC ACCTGCTGA GTGGTGGCG AATCTGTCA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 GAGTTTTC CTTCCAGGC TCGCAACAT GTGGTGGCG AATCTGTCA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
 GAGGCGGCA CAGTCCCGG TTTGCTGAT TGAACATCT GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA CCGTAAAGA
 CTTCCCGGT CAGTCCCGG TTTGCTGAT TGAACATCT GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA CCGTAAAGA
 1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
 GTTTCGACT GGCACGGCC AGACCGGCG TCTGCGGCT TCGTATGTC GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 CAGAAAGTA CCGTCCCGG CCGTCCCGG TCTGCGGCT TCGTATGTC GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
 TGTACCGCG TCTGCGGCG GGAACATGA AATCTGTCA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA CCGTAAAGA
 AATCTGTCA CCGTCCCGG CCGTCCCGG TCTGCGGCT TCGTATGTC GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
 GGTGCGGCA GAGTCCCGG TTTGCTGAT TGAACATCT GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA CCGTAAAGA
 CTTCCCGGT CAGTCCCGG TTTGCTGAT TGAACATCT GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA CCGTAAAGA
 1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
 ATAGTATTAC GAGGCGCAT CTGCTGATG GAGCGGCG TCGTATGTC GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 TATCATTAAG CTTCCCGGT CCGTCCCGG TCTGCGGCT TCGTATGTC GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
 TGGAGACGTT TCGTATGTC CAGTCCCGG TCTGCGGCT TCGTATGTC GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 ACCTCTGCA CAGTCCCGG TCTGCGGCT TCTGCGGCT TCGTATGTC GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
 GCAATTCGTT CAGAACGTA TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT
 CGTTTACCA GGTCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT
 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
 B GTTGGCGCG CCGTCCCGG ACGGGAAGG GGTGCGCTG TCGTATGTC GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 CAAACCGCG CCGTCCCGG TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT TCGCTGTGAT
 2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
 AGGAAAGGA GAACGAGGC GTATGAGCG CCGTGTGAG GCGAACGTA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 TCTTTTCTT CTTGCGTGG CATTACCGG CCGTGTGAG GCGAACGTA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
 CACCAAGCA GTTGGGTTG TAAACATCG GGTGTAGAG GCGAACGTA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 GTGGGTTCT CTTTGGGTTG TAAACATCG GGTGTAGAG GCGAACGTA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
 CCCCCGAGG AAAAAAGAA ACCTGCGTTA CCGTGTGAG GCGAACGTA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 GGGCGCTTC TTTTCTTTT TGGAGCAAT GCGCTGTGAG GCGAACGTA GCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
 AGAAGAGCA GAGGCTGAG CTCTGCGCG GAGTCCCGG AAAAAACCC CCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 TCTTCTGCT CTTCTGACT CTTCTGACT CTTCTGACT CTTCTGACT CTTCTGACT CTTCTGACT CTTCTGACT CTTCTGACT CTTCTGACT
 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
 AATCCAGTC AGCGCGCGT GCGCGCGCG GCGCGCGCG CCGCGCGCG CCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 TGTGGGAGC TCGCGCGCA GCGCGCGCG GCGCGCGCG CCGCGCGCG CCGCGCGCG ATAGAAATC CCGTAAAGA CCGTAAAGA CCGTAAAGA
 2510 2520
 C GTACCTCAGT CACCTCTTC
 C CATGAGTCA GTGAGAGG

FIGURE 20.1A–C. Nucleotide sequence of a region between coordinates 59.5 and 66.4 on the Ad2 genome. This sequence and the positions of strategic sequences were determined by Akusjärvi *et al.* (1981) (nucleotides 1–1164) and Kruijer *et al.* (1982) (nucleotides 858–2514).

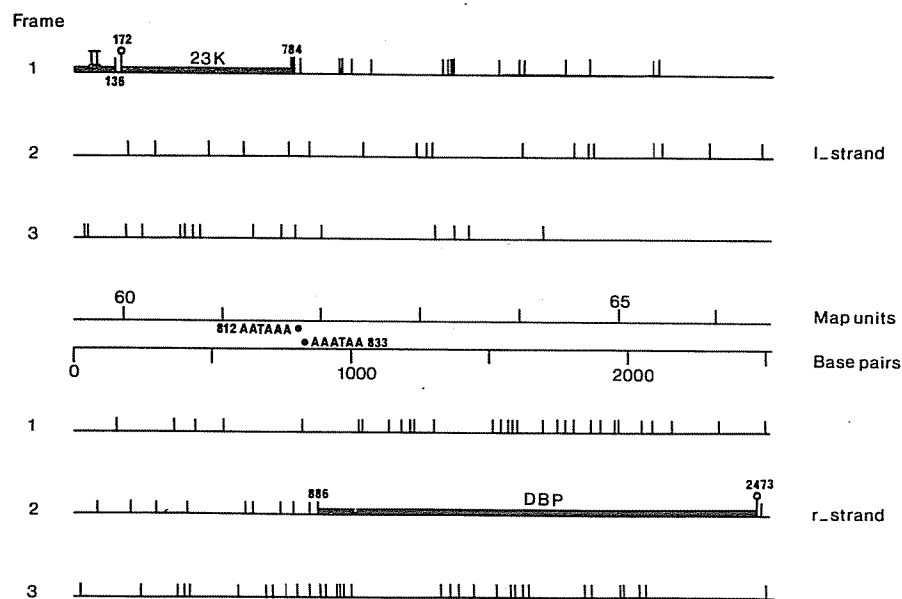


FIGURE 20.2. Structural organization of a region between coordinates 59.5 and 66.4 on the Ad2 genome. This map is derived from the nucleotide sequence in Fig. 20.1. For explanation of the symbols, see the Fig. 3 caption [Section VII].

10	20	30	40	50	60	70	80	90	100
5'-TTAAGACGGG	CGGTGGACGA	CACCGCAAGG	ATCGCTGAAT	CACGGGTAAT	TCATGGCACT	TACGGGAGGC	GGCGAAACCC	CAGTGACGAT	GGAAAGCTC
110	120	130	140	150	160	170	180	190	200
GATCGTTTGA	TGGAACGGAT	GGTGAAGCTG	TAGTACCTTC	TGCACCTGCC	ACTGCCGGAT	GACCTCACAG	TGACAGGAC	GTTGGATACG	TGGGGCTGG
210	220	230	240	250	260	270	280	290	300
CGAGGGACCA	GACGTTAAGC	GTTGACGAAT	CGCTTTCAGT	TTAATAGCCA	TGGAACCTCG	ACGTCCCAAG	GACGCACTG	CTTTTCAGGC	GCCGAGGCC
310	320	330	340	350	360	370	380	390	400
CAACTTTTGA	TGAGGCCCGC	ACACTGCGAG	CCGAATGGAA	GGCTTTAAAC	ATGGACTCCT	GATGGTCCGG	GTGCTCTAAT	CCAAGATGCT	TCGTGGTAGG
410	420	430	440	450	460	470	480	490	500
GGGGGGGGTT	TACGGCTCGA	ATGGCGGAGC	CAGTAATGGC	TCCCGGTGTA	GGAAACGGTT	AACGTTCGGT	AGTTGTTTCC	GGCGGTCTCT	AAAGACGAT
510	520	530	540	550	560	570	580	590	600
CTTTCCTGCG	CCGCCCAATG	GACCTGGGGG	TCAGGCCGCT	CCTCGAGTTG	GTTTACGGGG	CGGGCGCGGT	CGGATAGTTC	GTGCGGCCCC	GGGAACGAAG
610	620	630	640	650	660	670	680	690	700
GGTCTTACCG	TGGGTTTTTC	TYCGACGTGC	ACGGCGCGCG	CGGTGGGTGC	CTGCTCTTCC	TTATGACCTT	GTGAGTCEGT	CTCCTCCAAA	ACCTGCTCTT
710	720	730	740	750	760	770	780	790	800
CCTCCTCTAC	TACCTTCTGA	CCCTGTGGGA	TCTGTTCGGA	AGGCTCCGGC	TCTCTCACAG	TCTGCTTTGT	GGCAGTGGGA	GCCAGCGTAA	GGGACCGCGC
810	820	830	840	850	860	870	880	890	900
CGCGGGGCTC	TTAACCCTTG	GCAAGGGTGC	TAGCGATGTT	GGAGGCGAGG	AGTCCCGGGC	GGCGGTGACC	GACAAGCGGC	TGGGTTGGCA	TCTACCTGCT
910	920	930	940	950	960	970	980	990	1000
GGTGACCTTG	GTCCCGGCGA	TYCAGATTGC	TCGGCGCGCG	CAATCGGGTT	CTCGTTGTTG	TCGGGGTTCC	GATGGCGAGC	ACCGCGCGCG	TGTTCTTTGG
A	CACTGGAAC	CAGGGCGGCT	AAGCTAAGC	AGCGCGCGCC	GTTAGCCCAA	GAGCAACAAC	AGCGCCAAAG	CTACCCCTCG	TGGCGCGGCG

FIGURE 21A-K. Nucleotide sequence of a region between coordinates 70.7 and 100.0 on the Ad2 genome. This sequence was established by Galibert *et al.* (1979), Hérisse *et al.* (1980), and Hérisse and Galibert (1981). Short sequences were also determined by Zain *et al.* (1979a,b), Zain and Roberts (1979), Baker and Ziff (1980, 1981), Arrand and Roberts (1979), and Shinagawa *et al.* (1980). The region between 89.5 and 100 was also determined by Gingeras *et al.* (1982).

1010 GTATCAACGA CATATCTGCT	1020 ACGAGCTTC TGTCTGCAAG	1030 TGACACCC ACTGTGGGG	1040 GTGTAGAGG CAACATCTCC	1050 AAGCGGGCG TTCGCCGCC	1060 GGAAGAAGA GCTTTCTCT	1070 GATGGTAGT CTACCATCAC	1080 CCGCACCGA GGCGTGGCT	1090 AGGGGGCATT TCCCCGTAA	1100 GTAGCACGTA CATCTGCTAT
<i>start ORF 2</i>									
1110 ATGATGGCAG ACTACCGTC	1120 TAGAGATGTC ATCTCTACAG	1130 GGGGATGACG CCCCCTACTG	1140 TGGCGCCCGT ACCGCGCGCA	1150 CGCCCTCGCC GGCGCAGCG	1160 GTGCTGTGTC CAGCAACAGC	1170 TGCCCAAGT AGCGGTACCA	1180 GTCTTCGTT CAGAAGCAAA	1190 CCGCTGGCT GGCGACCGGA	1200 ATCTCTGTA TAGCAAGACT
<i>stop ORF 1</i>									
1210 GACTGTCTTC CTGCAAAAG	1220 GGTCTCTTAG CCAAGAAATC	1230 GTGTGCGCGC CACAGCGGG	1240 CGTGTGCTC GCAGCAGCAG	1250 CTCTCTCTC GAGGAGGAGC	1260 CGACGACAGC GCTGCTCTG	1270 CGCGGTTTSE GGCGCAACG	1280 TTGGGCATAG AACCGGTATC	1290 CTGGCGCTC GACCGCGGAG	1300 GAATCTTAT CTTAGAAATA
<i>splice E2A leader</i>									
1310 CCTAAAGAGG GGATTCTTTC	1320 GTGAGACATA CACTCTGTAT	1330 CGATATAAAG GCTATATTTT	1340 TTGTTTCTGTC ACCAAGCAG	1350 CGCGCTTCTT GGCGCAAGAA	1360 GTCTCTGACT CAAGAGCTGA	1370 TTTATTTTTT AAATATAAAA	1380 GTCCAGAGAC CAGGTCTCTG	1390 GGGAGGAGT CGCTCCCTCA	1400 GGGCGTCGAC CGCGAGCTG
<i>start E2A RNA early</i>									
1410 GGACATAGTG CCTGTATCAC	1420 TTTTGCTTTC AAAAGCGAAG	1430 TAGTGAAGC ATCAGCTTCG	1440 CGCGTGGCAG GGCAGCGCTG	1450 CTTCTGCGCC GAAGACGCGG	1460 TCCGAGAGAA AGGCTCTCTT	1470 GTGCTTTATG CAGCAATATC	1480 ACCGCGGACT TGCGCGCTGA	1490 GAGAACTTCT CTCTTAAGGA	1500 GATCAAAAGC CTAGTTTTCG
<i>stop ORF 2</i>									
1510 CGGGAAGAGG GGCTTTTCTC	1520 TTTAAATTCG AAATTTAAGC	1530 CGCTTTTGTAT GGCAAACTA	1540 GCATAGAGG GCTCATCTCC	1550 TCCCGCTGT AGCGCGCACA	1560 GGCGCGGCT CCCGCGCGCA	1570 CGTGGACAGC GCACCTGTGC	1580 AGTCCGGTA TCAGCGCCAT	1590 ATACGCTTC TATGAGCAAG	1600 CTTTAAGGCT GAAATTCCTCA
<i>start protein pVIII</i>									
1610 CGGGATGATA CGCCCTACAT	1620 CACTCAATG GTGGAGTTAC	1630 GTGCGTGT GCGCGCAATG	1640 ACCTGAACG TGGGACTTGC	1650 CCGACTCGA GGCTGGAGCT	1660 CGGGTCTGA GCCCAAGACT	1670 TGAGTTGGG ACTCAACCG	1680 TTATTTGATG AATTAATCTC	1690 TACTCGGCC ATGAGCGCGG	1700 CTGGGGTGA GACCCCAAT
1710 CTATAGGCC GATATCCCGG	1720 CAGTTGCCCT GTEAACGGA	1730 AGCGCGGGT TCCCGCGCCA	1740 GGCTTGGCT CCGAACGGA	1750 TAAGAGGAGC ATTCTCTCTG	1760 TTGTCGCCG AACAGCGCGG	1770 ATAATGGTGG TATTACCAAC	1780 TGTGGAGCAT ACACCTCGTA	1790 TATTGGAATT ATAACTTAA	1800 AGGGGCAATCA TCCCGGTAGT
1810 ACGGGCGGAC TGCGCGCGTG	1820 GGGACACAT CCCTGGGTGA	1830 GGTCTTTTCA CCAGGAAGAT	1840 GGCGGAGGCT CCGCTCTCCA	1850 GGTGACACCA CCACTGTGGT	1860 TGAAGGGTCT ACTTCCCGGA	1870 CTCGCGGTC GACCGCCAGG	1880 GGCTTCAAGT CCGAACTTGA	1890 CTACTGATTG GATGACTAAC	1900 AGTCCCGCGG TCAGCGCGCG
1910 TGAACGCGCC B AGCTTGGCGG	1920 CGCGAAGCA CGGCTTTCGT	1930 GTGTGCGCAC CACAGGCTGC	1940 CCAGCGGGCC GGTCCGCGCG	1950 CGTCCCATAT GCAGGGTATA	1960 TGAGTGGACT AAATCAGAGG	1970 TTTAGTCTCC AAATCAGAGG	1980 CGCTCCATAA GGGAGGTATT	1990 GTGCGATTGC CAGCTCAACG	2000 TGCTCAGCCA ACGAGTCCGT
<i>start E3 RNA</i>									
2010 CTCGAGGAGA GACCTCTCT	2020 GAACCAAGGG CTTGGTCTCC	2030 CAGCGCTGCC GTCGCGAGG	2040 CTGTAAAGTC GACATTTTCAG	2050 TAGCCGCGCC ATCGCGCGCG	2060 GACCGCGGAG CTGGCGCTC	2070 AAGTAATGC TTCATTTAGC	2080 GGGCGCATCC CCCCGTCAAG	2090 GCTAGGATTG CGATCTTAAC	2100 AGAGCTCTCG TCTCGAGCC
2110 AGCAGGAGGC TCGTCTCTCG	2120 TCGGCGCGAG AGCGCGCGCT	2130 GCCTCCGTAA CGGAGCGATT	2140 CCTTGAGATG GGAATCTTAC	2150 TAAATAACT AATTTATTGA	2160 CCTCAAGCAC GGAGTCTGCT	2170 GGAAGCAAA CTTCTGGGTT	2180 TGAAGTTGGG ACTTCAACCC	2190 GAAAGACCT CTTTTCTGGA	2200 GGAGGCGCGG CCTCCGCGCG
2210 TGATGGGCTC ACTACCGGGA	2220 GGTCAATATA CCAGTATTAT	2230 GGTTTGAAC CCCACTTTTG	2240 TGCGGCACTT AGACTCGGCG	2250 TCTGAGCCG AGACTCGGCG	2260 CTGCGGATGC GACGGCTACG	2270 TGACTTACTG ACTGAATGAC	2280 GTCACCTCTC CAGTGGAGAG	2290 CGTCTCGCTG CGAGAGCGAC	2300 ACCGGAGCTG TGGCGCTGAC
<i>splice 1 x leader</i>									
2310 TGTGAGGCTG ACACCTCGAC	2320 GTGACGCGCG CACTGCGCGC	2330 CGGTGTTCAC GGCAAGATG	2340 GAAACGGGCG CTTTGCGCGC	2350 CGGAGGCCAC GGCTCGGCTG	2360 TCAAAACAA AGTTTGTGTA	2370 GAACTTAACT CTTTGAATTT	2380 GGGCTTCTCG CCCGAAGAGC	2390 TATAGTCCCG ATATCGAGGG	2400 GGGCGCGCGT CCCGCGCGAC
<i>splice 2 x leader</i>									
2410 CGGAGCGCGC GGCTCTCGCG	2420 AGTGGTGGGT TCACACCECA	2430 CCATCTCGAA GGTAGAGCTT	2440 TGTGCATCG ACACGTAGCC	2450 ACTAAGCCCT TGATTCCGGA	2460 CAAATGGTTC GTTTACCAAG	2470 GCGGGGGAGC CGCCCTCTCG	2480 ATCACCTCGC TAGTGGAGCG	2490 CCTCGCGCCA GGAGCGGGGT	2500 GGGACACAA CCCTGTGTTC
2510 ACTGGCACCA TGACCGTGGT	2520 AACGTTGACA TTGCAACTGT	2530 GGATTGGGAC CCTAACCTTG	2540 CTAATGTAGT GATTACATCA	2550 TCTAGAAACA AGATCTTTGT	2560 ACAGTAGAGA TGCTGAGTA	2570 CACGACTCAT GTCTGAGTA	2580 ATTATTTATG TAATAAATAT	2590 TCTTTAATCT AGAAATTAGA	2600 TAGATGACCC ATCTACTGGG
2610 CGAGGACAGC GCTCTGTGCG	2620 GGTAGACAC CCATCTCTGT	2630 TTGCGGTGGC AACGCCACCG	2640 AAAAATGGGT TTTTTACCCTA	2650 GGGTTTCTGT CCCAAGCAG	2660 TGGTTTCTGT ACCAAGCAAA	2670 TGGAGTGGAG ACCTCACCTC	2680 GCCAAACGTG CGTTTTCAC	2690 TTGCGCGCGT AAGCGCGCCA	2700 TATTCATGGA ATAAGTACCT
2710 ATGGACCATG TACCTGGTAC	2720 AAATTGCCGA TTTAACGGCT	2730 GAAGTAAACA CTTCAATTTGT	2740 TTAAATGTTG AATTTACAA	2750 TCAAGGTGCG AGTTTCCAGC	2760 CTCTGCTTCA GAGACGAAGT	2770 TTCAACGGT AAGTTTGCCA	2780 GTGTTGGAA CACAACCTTC	2790 AGCCGAAGTT TCGGCTTCAA	2800 GATGTGGCAG CTACACCTGC
<i>splice y leader</i>									
2810 TTCTTTTGT AAGAAACAA	2820 GGTGGTGGT CCACCAACAC	2830 GTGGAGGAG CACCTCTCTC	2840 TGGACGGCCG ACCTGCGCGG	2850 TTGCATGCTC AAGCTAGCAG	2860 ACCGAGTGGC TGCGTCAACG	2870 CAACGACCGG GTGCTGCGCG	2880 GGTGTGGATG CCACACCTAC	2890 TCGGACTGCG AGCTGAGCGC	2900 ATTGGTCTGT TAACACAGACA
2910 AATGAGGDTA CTTACTCCCAT	2920 AAAAGTTTTT TTTTTCAAAA	2930 GTCTCCCACT CAGGAGGTGA	2940 CGAGTTGAGG GCTCAACCTC	2950 GCCTTGAAGT CGGAACCTCAG	2960 CAGTTTCTTC GTCAAAAAGG	2970 GTAAACCGCC CATTTTGGCG	2980 CCACGACCTT GGTGTGGGA	2990 AAAAATTAA TTTTTAAAT	3000 TTCTATACT AAGTATATGA

FIGURE 21 (Continued)

3010 CGTAAAGTTC GCAATTCAAG	3020 ATTGAGATGT TAACCTACAA	3030 TCGAACAGAT AGCTTGCTTA	3040 TAAAAAGACC ATTTTCTGCG	3050 TTAACCCAG AATTGGGGTC	3060 CCCCAATAGG GGGGTTATCC	3070 AATGAGAAC TTACTCTGT	3080 TTAAGACAAA AATTCCTGTT	3090 TAAGAATATG ATTCCTATAC	3100 ATCGTAGAGA TAGCACTCT
3110 CACGGAATCC GTGCTTAAAG	3120 CAACGGCGGA GTTGCGCGCT	3130 CGACGTGCGT GCTGCACGCA	3140 GCAACATGCG CGTTTGTACC	3150 ATACACATCG TATTGTGACG	3160 AAAAATTGCG TTTTTAAAGC	3170 GACCCCGGTT CTGGGGGCAA	3180 GTAGGTTCTA CATCCAGAT	3190 CTCCATGTAC GAGGTACATG	3200 TAAATCCGA ATTTAGGCT
3210 ACGACGGGGA TGCTGCGCT	3220 ACGCGCTCAG TGCGGCGATC	3230 ACGTGCGGAC TGCAGCGCTG	3240 GGTTTTTCCA CCAAAAAGGT	3250 ACTCAAAATC TGAGTTTAAAG	3260 CTTGGTCCAA GAACACGCTT	3270 CGTTACAAATG GCAATGTATC	3280 TAAATTTAGT ATTTAAATCA	3290 CTTCGATTAC GAAGCTAATG	3300 TTACGTGATG AATGCACATC
3310 AGAATATTTT TCTATATAAA	3320 ACGTGGTGTG TGACACCAAG	3330 TTGTACTTTT AACATGAAAA	3340 CGAATAATTA GCTTATTATT	3350 GCGGTGTGTC CGCCACAAAG	3360 TGTTTTTAAC ACAAAATTGG	3370 GTTCAATGCA CAAGTATGCT	3380 CATATACGAT GTATATCTTA	3390 AAACCGTCGG TTTGGCAGCC	3400 TCCACTGTGA AGGTGACACT
3410 TTGCTGTAT AACGACTATA	3420 TACAGTGTCA ATGTCACAGT	3430 GAAGGTTCCT CTTCCAAAGT	3440 CTTTTAGCAT GAAAATCGTA	3450 TTTGAAAAAT AAACTTTTAT	3460 CATATTTAAA GTATAAATTT	3470 GGTAAATATC CCATTTTATG	3480 TTTACACGCT AAATGTGCGA	3490 ATAATGGTAC TATTACCAATG	3500 ATGTACTCGT TACATGAGCA
3510 TTGTCATGTT AACGACTACA	3520 CAACACGGGG GTTGTGGGCC	3530 GGTGTGTTTA CCACAAAAGT	3540 CAATCTGTTT GTTTAGAGAA	3550 GTGACGTTGG CACTGGCACC	3560 AAACAAAGCT TTTTTGTCCA	3570 GGGAGAGGCA CCGCTCTGCT	3580 ATAATGTCCT TATTACAGCG	3590 GAAGGAAAG CTTGCTTTGG	3600 ATACATGTGA TATGTACTCT
3610 TGAATATAGG ACTTATATCT	3620 TTTATGTGTT AAATACAAAA	3630 CGTCTGCGTC GCACACGACG	3640 AAATAAATCA TTTTTATGAT	3650 CTTTTCTTTT GAAAAAGAAA	3660 ACGGAATCAA TGCCCTGATT	3670 AAGCGGAACG TTCCGCTTGC	3680 AACATAAGGG TTGTATATCC	3690 GACCTGTAAA CTGGACAAAT	3700 ATGAGATACA TACTATATGT
3710 CCCTATACGA GGGATATGCT	3720 GGTCCGCGCC CCAGGCGGGC	3730 TTCTAATATG AAGATTATAC	3740 GGTGTGGGAA CCACAACCTT	3750 GTTTAGTTTG CAAAATCAAC	3760 AAAGGACCTG TTTCTTGGAC	3770 CAATCGGGCA GTTAGCGCCT	3780 CTAAGACAGG GATTTCGCGC	3790 TCGGGAGAGT AGCGCTTGCA	3800 GACGTGTTAA CTGCAAAATT
3810 CTAGTTTGGG GATCAAAACC	3820 TGAAGTCTGA AGCTTACGCT	3830 ACGGACGAGG TGCTGCTCTC	3840 TCTCTACTGG AGAGTGGACC	3850 CCGAGTTGGT GGCTCAACCA	3860 AGCGCGGGTG TCGCGCCGAC	3870 TTGCTGTGTA AACGGACTAT	3880 GCGTTGTGGT GGCAACACCA	3890 GACGATGGCC CTGCTACCGG	3900 TGATTGTAGA ACTAAATCT
3910 CGGGATTAA D GCGCTAAAT	3920 ATGGGGTCCA TACCCCAAGT	3930 AGTACGGAAA *CATGCTTTT	3940 CAGTTACTGA GTCAATGACT	3950 CCCCCTGCAA GGGCGAGCTT	3960 CCTGTACACC GGACATGTGG	3970 ACCAAAAGGT TGGTTTTCCA	3980 -ATCCGGAATA TAGCGCTTAT	3990 CAACCAAAAG GTTTGTGTCG	4000 GAATATATAT CTTATTATTA
stop URF 13									
4010 ACACCGAATA TGTGCTTAT	4020 AACAAAGGAT TGTGTGCTTA	4030 TTCGGCTCTG AAGCGCAGAC	4040 CGCGGTCTGG GGCGCAGACC	4050 GGGGTAGATA CCCCATCTAT	4060 TCCGGATAGT AGGCCATATCA	4070 AACACGAGTT TGTGTGCTCA	4080 GGGTGTGTTA CCCAACAAAT	4090 CTTTTTTAAG GAAAAATTC	4100 TATCTAACCT ATAGATTGGA
4110 GCCAGACTTT CGGTCTGAAT	4120 GGTCAAGAG CCATGTCTCT	4130 AAGAAATGTT TTCCTTTTACA	4140 CATACTGTAT GTATGATTTAA	4150 TACTCTGTAC ATGAGACATG	4160 TAAGGAGCTC ATTCCTCGAG	4170 AAGAAATATA TTCTTATATT	4180 TAACCTGGAA ATTGACCTTT	4190 CAAGCGGAAA GTTGCGCTTT	4200 AGACACGAC TCTGTGCGTG
4210 GAGATGTAA CTCTACATTG	4220 CGCGCGCCAG GGCGCGGTCG	4230 GAGTGTAGCT CTCACATCGA	4240 TCATCTAACG AGTAGATTGG	4250 TAGGGTGGAA ATCCCACTTT	4260 AGTGTCAAAAT TCACAGTTTA	4270 GGACGAAATG CGATTGTGCA	4280 CCTAAACAGT GGATTGTGCA	4290 GGGAATAGGA CCCTTATGCT	4300 GTAGACGTGG CATCTGCGAG
4310 GAGCAGTGAC CTGCTCACTG	4320 ATCAGTAGCG TAGTCATCGC	4330 GAAGTAAGTC CTTCATTTCAG	4340 AAGTAATCTG TTCATTGACT	4350 CCCCAACACA GGGTTTGTGT	4360 CGCGTAACGC GCGATTGCGG	4370 ATGGAGTCCG TACTCTACGG	4380 TGTAGGCGCT ACCATCCGCA	4390 TATGTCTCTG ATACAGAGAC	4400 TCTGTATATC AGGCATATAG
4410 GACTAGAAAG CTGATCTCT	4420 GTCTTAAGAA CAGAAATCTT	4430 ATTAATACTT TAATTAAGAA	4440 TGCTCTCACG ACGGAGTGTG	4450 TAAAAACAAA ATTTTGTGTT	4460 ACGACTAAAA TGCTGATTTT	4470 AACCGGGGAT TTGCGGCCCTA	4480 GGACACGAAA CCTGTGCTTT	4490 CGAGGGTTTG GCTCCCAAC	4500 GAGTCCGCGA CTCAGCGCCT
4510 GGGTTTTCTG CCCAAAAGAC	4520 TATAAAGGAC ATATTTCCTG	4530 GTCTAAGTGA CAGATTTCAC	4540 GTTTATAGCT CAAAATATGGA	4550 TGTAAGGGTC ACATTCCGAG	4560 GACGATGTTG CTGCTACAAC	4570 TTTGTCTCGC AAACAGAGCG	4580 TAJACAGCTT ATTTGTGAGA	4590 TCGGACCAAT AGCCTGGTTA	4600 ATGCGGTAGT TACCGCATCA
4610 AGAGACAGTA TCTCTGTCAT	4620 CCAAAAACCG GGTTTTTTTG	4630 TCATGGTAAA AGTACCATTT	4640 AACGGGATCG TTGCCCTAGC	4650 GTATATAGGT CATATATCCA	4660 ATGGAATGCT TACCTTGACA	4670 AACCGACCTT TTGGGTGGAA	4680 ACGGTATCTA TGCCATAGAT	4690 CGGTACTTGG GCCATGAACC	4700 TGGGATGAAA ACCTTACTTT
4710 GGGTACGCGG CCAGTGCCCT	4720 CGACAGTATG GCTGTCTATC	4730 GTGACGTTGT CACTGCAACA	4740 CCAATAACGG GGTTATTGCC	4750 GGTTAGTTAG CCAATCAATC	4760 TCGGAGCGGG AGCCTCGCCC	4770 GGGAAGAGGG CCCTTCTCCC	4780 TGGGGGTGAC ACCCCACTG	4790 TCTAATCGAT AGATTAGCTA	4800 GAAATTAAC CTTTAATTG
4810 TGTCCACCTC ACAGTGGAG	4820 TACTGACTTA ATGAGTGGAT	4830 GAGATCTADA CTCTAGACTT	4840 TCTTAACCTA AGAATTGGAT	4850 CCTTAATTGT GGAATTAACTA	4860 GGCTTGTGCG CCGACAGCGG	4870 GGATGATCTT CCTCATAGAA	4880 TCEGCGTTCC AGCGGCAAGG	4890 GCCCGAGGCT CGCGCTCCGA	4900 CGCTCTTGGG GCGAGAACGG
4910 GATTTTGTTC E CTAAGAACAG	4920 TTCAACTTCT AAGTTGAAGA	4930 GTACCAATTG CATGGTTAAC	4940 GATGTGGTCA CTACACAGT	4950 CATTTTCTCC GTAAAGAGG	4960 ATAGAAAAAC TATCTTTTGT	4970 CACCAGTTGG GTGCTCAAGC	4980 TCGGGTTTGA AGGCCAAACT	4990 ATGGATGCTT TACCTACGAA	5000 TTTTGTGAT AAAACCACTA
splice 2' leader									

FIGURE 21 (Continued)

5010 GGCCGTTGGC CGGCAACCG	5020 GGAGTCGATG CCTCAGCTAC	5030 TTCGATGGGT AAGTACCCA	5040 GGTCCGGATT CCGAGGCCA	5050 TTTGGACCAC AAACTGGTG	5060 GAATACCACC CTTATGGTGG	5070 CTCTTTTGGC GAGAAAAAC	5080 ATAGTGGCAG TATCACCCTC	5090 TGGGTCTGTA ACCCAGCACT	5100 GCCGCTTTTG CGGCAGAAAC
5110 TCTCCCGAGC AGAGGGTGGC	5120 GACGTGAAGG CTGCACCTTC	5130 GGATAGTCCC CCTATCAGGG	5140 AGGTCTCCTE TCCAGAGGAC	5150 GAGACGTGAG CTCTGCACCTC	5160 AATAATTTTG TTATTAANAAC	5170 GTACACACCA CATGTGTGGT	5180 TAATCTCTAG ATTAGAGATC	5190 AATAAGGTAA TTATTCATT	5200 GTGATTGTA CAACACACT
5210 TTTGTGTGTT AAACACACAA	stop URF 18 ATTIAATGAA TAAATTAATT	5220 TGAATTTTAA ACTTAAATC	5230 TCAGTGGTTT AGTCAGCAAA	5240 AGAAACAGGT TCTTTGTCCA	5250 CGAATAAGTC GCTTATTCAG	5260 GTAGTGGAGG CATCACCCTC	5270 AAAGGAAGGA TTTCTTCTCT	5280 GGGTGAGAC CCCAACTCTG	5290 CATAGAGTCG GTATCTCAGC
5310 GCGGAAAATC CGCTTTTAG	5320 GACGTTTGAA CTGCAAACTT	5330 AGAGGTTTCA TCTCAAAAGT	5340 AATTTACCCCT TAAATGGGA	5350 ACAGTTTAAAG TGTCAAATTC	5360 GAGTACAAAG CTCATGTCTT	5370 ACAGGGAGGC TGTCCTCCCG	5380 GTGGGTGATA CACCACACTAT	5390 GAAGTATAAC CTTCATATTG	5400 AACCTCTACT TTGCGAGTGA
5410 TTCCCGCGTC AACGCCCCAG	5420 TGGCAGACTT ACCGTCTGAA	5430 CTGTGAAAGT GACACCTTCA	5440 TGGGGCACAT ACCCCTGTGA	5450 AGGTATAGTG TCCATATGAC	5460 TGCTTTTGGC ACAGAAACCG	5470 CCGGAGGTTG TGTGCCCTTT	5480 ACACGGGAAA TGTCCTCTCT	5490 GAATGGGGAG CTTACCCTTC	5500 GTAAACAAAG CATTTGTTTC
5510 TGGGTACCA ACCAATAGGT	5520 AAGGTTCTTT TTCCAAGAAA	5530 CAGGGGACCC GTCCCTCGTG	5540 TCAAGAGAGA AGTCTCTCTT	5550 GATGCCGACA CTACCCCTCT	5560 GCTTGCAGAA CCGAAGCTTT	5570 CTGTGGAGGG GACACACTCC	5580 GTCCCTTACG CAGCGCATGC	5590 AACCGGAATT TTGCGCTTAA	5600 TTACCGCTCC AATGGCGAGC
5610 CCAGATGGG GGTCTTACCC	5620 ATCTGTTCGG TAGACAAGGC	5630 GCCTTTGGAG CGGAACCTCT	5640 TGGAGGGTTT ACCTCCCAAA	5650 TACATTGGTG ATGTAAACAC	5660 ACAATGAGTC TGTTACTCAG	5670 GGTGAATTTT CCACTTAAAA	5680 TTTGTTCAG AAACAAAGTC	5690 TTTGTATTCA AAACATAAGT	5700 AACCTGTGGA TTGGACACTT
5710 GGCGTGGTGA CCGCACCACT	5720 ATGTTAATGG TACAATTACC	5730 AGTCCGCGGG TCAAGCGCCC	5740 ATTGTACCGT TAAAGTGGC	5750 TTGTGTGCGA AACCAACGCT	5760 GGAGACTATC CCTCTGATAG	5770 AATGATCGCC TTACTAGCGG	5780 GCGAGACTCG CGCTCTTAGC	5790 CATGTACAGT GTACAGTCACT	5800 TTCCGGGTGA AAGCCCACT
5810 CTGGCACGTT GACCTGTCGA	5820 CTGAGGTTTG GACTCCAAAC	5830 ATTGCTAAGC TAAAGATTGG	5840 ATGATTCCCC TACTAAAGGG	5850 GGGTAAATGC ACCAATACAG	5860 ACAGTCTACC TGTCAGATGG	5870 TTTCCATCGG AAAGCTAGCC	5880 GACGTTTGTA CTGCAACATC	5890 GTCCGGGGGA CAGCCGCCCT	5900 GAGACCGTCA CTCTGGCAGT
5910 CTGTGCGTGT FGACAGCGACA	5920 GGGAATGACA CCCTTACTGT	5930 TTGACGTAGT AACTGCATCA	5940 GGGGCGGATT CCCCCGCTAA	5950 GATGACGGTG CTACTGCCAC	5960 CCCATCGAAC GGGTAGCTTG	5970 CCGTAATGTT GGCATTAAAC	5980 ACCTCTTAGG TGGAAATGCC	5990 ATAAATACAT TATTTATGTA	6000 TTATTACTCT AATAATGGAA
6010 TTTACCTCTTA AATATGAAT	6020 ATTTTATTCG TAAATAAAGC	6030 CCAGGAAACG GGCTCTTTGC	6040 TTCATCGTGT AAGTAGCACA	6050 TTTGAGGCTA AAACTCCGAT	6060 TGTGATTGTC ACACTAAGAC	6070 ATCAATGACC TAGTACTGCG	6080 TGGTCACAG ACCAAGGTGC	6090 TGGCAACTTG ACCGTGGAAC	6100 TTTTGAGGGA AAACTCCCT
6110 ATCTGTGTTT TAGACCAAA	6120 CAAGCTCCTC GTTGCAGGAG	6130 GATAACCAAT CTATTGGTTA	6140 ACTAAGTAGT TGATTCTACA	6150 TTGTTGTACC AACAACATGG	6160 TTTAATTTTG AAATTAANAAC	6170 CCCGCCACCG GGGCGGTGGC	6180 TACCGATATT ATGCGTATAA	6190 TATTGTTGAA ATAACAATCT	6200 CAATTAAAGT GTTAATCTTA
6210 CTACACCTAA GATGTGGATT	6220 TGGGTAAACT ACCCATTGTA	6230 ACGAGTTTGT TGCTCAAAAC	6240 TTTGATGCGAG AAACTAGCTC	6250 AATTTGACCC TTAAACTGGG	6260 CGTCCCTGGG CGAGGGACCC	6270 GACATATAT CTGTATATTA	6280 TACGTAGAGT ATGCACTCA	6290 ATTGAACCTG TAACCTGGAC	6300 TATTTGATAT ATAAATCTTA
6310 TGCTCTCGGA ACAGAGGCTT	6320 TATGGAAAAA ATACCTTTTT	6330 TTACGTAGTT AATGCATCAA	6340 TGTTATGATT ACAATACTAA	6350 TTTTGACCTT AAAACCTGGA	6360 CAATCGTATT GTTAGCATAA	6370 TTTTTAGGTC AAAAATCCAG	6380 ACCTGATTTG TGGACTAAAC	6390 AAACTATTAT TTTGATATTA	6400 GACGGTATCG CTCCGATAGC
6410 ATATTTCAGT TATAAATGCA	6420 CCTTTCCCGC GGAAGGGTCT	6430 ACCTCAAACT TGGAGTTTGA	6440 ATGTTTGTGT TACAACACAA	6450 AGACTCAGAG TCTGAGTCTC	6460 GTCTATAGTT CAGATATCAA	6470 GGGTTATTTT CCCAATAAAA	6480 TGATTTTAAAC ACTAAAATTT	6490 CGAGACCGTA GCTCTGGCAT	6500 ACTAATGTTA TGATTACAA
6510 CTTTTCCGAC GAAACCGGTG	6520 GGTACTAATG CCATGATTAC	6530 ATTGAACTCT TAAACTTGGG	6540 CGCCCAAAAT GGCGGTTTTA	6550 CGAAACTGTT GCTTTGACAA	6560 GAGTCCCGGG CTCAGGGGCC	6570 TAATGTTATC ATTACAATAG	6580 CTTTGTTTTT GAAACAAAAA	6590 ACTACTGTTT TGATGACAAA	6600 GAATGGGACA CTTACCTCTG
6610 CCTGTTGGGG GGACAACCCC	6620 TCTGGGTAGA AGACCCATCT	6630 GGATTGACGT CCTAAGTCCA	6640 CTTAAGTAAAG GAATTCATTC	6650 TCTATTACTG AGATAATGAC	6660 ACGTTTAAAT TGCAAAATTTA	6670 GAAACCAAGA CTTTGCTTCT	6680 ATGTTTTACA TACAAAATGT	6690 CCCTCAGTTC GGGAGTCAAG	6700 ATGATCGATG TACTAGCTAC
6710 ACATCGACGA TGTAGCTGCT	6720 AACCAGACATA TTGGCTGTAT	6730 GACCTCTAGA CTGGAGATCT	6740 AAGTAGGTAC TTCATCCATG	6750 TGTCCTGGGC ACAGGCAKCC	6760 AACGTTTACA TTGCAAGTGT	6770 ATCATATAAG TAGTATATTC	6780 GAATCTAAAC CTTAGATTGG	6790 TGCTTTTGGC ACCAAAACGG	6800 ACAAGATTAC TGTTCTAATG
6810 CTCTTGAGGA GAGAACTCT	6820 GTGAATTTTT CACTTAAAAA	6830 TGAATAGACC ACATTAACTG	6840 TTGAAATCTT AACTTTAGAA	6850 TACCTTTGAG ATGGGAATCT	6860 TTGATTAGCT AACTAATGCA	6870 TTAGGTATGT AATCCATACA	6880 GTTTACGTCA CAAATGCAGT	6890 ACCTAAATAC TGGAATTATG	6900 GGATTGGAA CCTAACCTTC
6910 ATCGGATAGG TAGCCATATC	6920 TTTTTGGGTT AAAAACCCAA	6930 TCAGTTTGAC AGTCAAACTG	6940 GATTTTATTT CTAAAATAAA	6950 GTAACAGTCA CATTTGTCAGT	6960 GTTCAAATGA CAAGTTTACT	6970 ACGTAACACT TGCAATGGTA	6980 ATTTGTATTT TAAACCTAAA	6990 GGATACATAG CCTATGATAC	7000 AATGGTAATG TTACCATTAC

FIGURE 21 (Continued)

7010 TGAATTACCG ACTTAATGCG	7020 TGATCTCTTA ACTAGTAGAT	7030 GGTGTCTTTC CCACAGAAAC	7040 ATCGCTCCAT TAGCGAGGTA	7050 TCGTGAATGA AGCACTTACT	7060 GATACAGAAA CTATGTCTTT	7070 ATGTACAGGG TACATGGTCC	7080 ACCCTTTCAC TGGGAAATGC	7090 CTTTTATGTG GAAATATCAC	7100 GTGACTTTGA CACTGAAACT
7110 AAACGATGGT TTTGTACACA	7120 TGAGATGTG ACTCTTACAC	7130 GAAGAGGATG CTTCTCCTAC	7140 TAACGGGTCC ATTGCCAGGG	7150 TTATTTCTTA ATAAAGAAT	7160 GCACCTTGGAC CGTGAACCTG	7170 AACGTACAAAT TTGCATGTTA	7180 ACAAAGTTGC TGTTTCAACG	7190 ACAAATAAAA TGTTTATTTT	7200 AGTTAACGTC TCAATTGCGC
7210 TTTAAAGTT AAAAATTGCA	7220 CACTAAAGG GTCTTTTTC	7230 TAAGTCATCA ATTGAGTAGT	7240 TATGGGGGTG ATAGCCCEAC	7250 GTGGTGTATC CACCATATAG	7260 GAATATAACT CTTATATTGA	7270 AGTGGCATGG TCACCCATAC	7280 AATTAGTTTC TTAATCAAAC	7290 ATGTCTTGGG TCACAGAAAC	7300 GATCATAGGT CTAGTATTCA
7310 TGGACGGTGG ACCTGCCACC	7320 AGGAGGGGTT TCCCTCCCAA	7330 GTGTGTCTCA CACACAGAGT	7340 TGTGTACAGA ACACAGTCTT	7350 AAGAGGGGCC TTCTCCCEGG	7360 GACCGGAATT CTGGCTTTAA	7370 TTTCTGTAGTA AAAGCATCAT	7380 TAGTACCCAT ATCATGGGTA	7390 TGTCTGTATA ACAGACATAT	7400 AGAATCCACA TCTTAGGTGT
7410 ATATAAGGTG TATATTCCAC	7420 TGGCAAAAGG ACGGTTTCTT	7430 CAGCTCGGTT GTCCAGCCAA	7440 TGGGAGTAGT ACGCTCATCA	7450 CACTATAATT GTGATATTAA	7460 ATTTGAGGGG TAAACTCCCC	7470 CCCGTCGAGC GGGCAGCTCG	7480 GAATTCAGGT CTTAAAGTCA	7490 ACAGGCGACG TGTCCGTGTC	7500 GTGAGCACT CAGCTGCTGA
7510 CGGTGTCCGA GCCACAGGCT	7520 CGACGAGGTT GCTGTCCAC	7530 AAGCCCAAGC TTGCGTTGTC	7540 AGTTGCCCGC TCAACGGGCG	7550 CGGTTCCTCT GCGAAGGGGA	7560 TCAGGTGCGG AGTCCACGCC	7570 ATGTACCCCC TACATGGGGG	7580 ATCTCAGTAT TAGAGTCATA	7590 TAGCAGGTAG ATGCTGCATC	7600 TCTTATCCCG AGGATAGGCG
7610 CCACACGACG GGTGGTGGTG	7620 GTGTCGCGCG CGAATAAATC	7630 GCTTATTGGA CGAGCGGCGG	7640 CGACGGGCGG GCTGCCGCTC	7650 GGCGAGGCGG CCGCTCCGTC	7660 GACGTCTCTTA CTGACGAAT	7670 TGTGTACCGG ACAACTAGGC	7680 TCACAGAGGG AGTGGTCTCC	7690 AGTGCCTACT TCAGCGATGA	7700 AAGCTGGGCG TTCCGACCGG
7710 GGCGTGGTAC CCGACGATGT	7720 TCTCGCGAAC AGAGCGCTTG	7730 AGGAGGCCCG TCTCTCGGCG	7740 TGTGTCGCGG ACAGCAAGCG	7750 TGGGACTAGA ACCTGTATCT	7760 GTGAATTGAG CACTTAATTC	7770 TGTGTCTCAAT AGCAGAGTAA	7780 GAGCTGCTGT CTGACGACAA	7790 CGTGGTGTTA GCACCAATAT	7800 TAAAGAGTTT ATTTGTTCAA
7810 TAGGGTGTGA ATCCACAGCT	7820 CGTTCGCGGA GCAAGGCGCT	7830 CATAGGTTTC GTATCCAAAG	7840 GAGTACCGCC CTCATGGGCG	7850 CCTGGTGTCT GGACACAGAA	7860 TGGGTGCACC ACCCAGCTGG	7870 GGTAGTATGG CCATCATACC	7880 TGTTCCGCTC ACAAGCCGAG	7890 CATCTAATTC GTAGATTAAAG	7900 ACCGCTGGGG TGGCGACCCC
7910 AGTATTGTGT ATCATAAACAC	7920 CGACCTGTGA GCTGGACATA	7930 TTGTAAAGTA AACATTACCT	7940 GAAACAGCTA CTTTGGCAT	7950 CAACATTAAAG GTTGTAATTC	7960 TGGTGGAGGG ACCACTTCCC	7970 CCATGCTATA GGTACCATAT	7980 TTTGGAGACT AAACCTCTGA	7990 AATTTGTACC TTAAACATGG	8000 GGCGTAGGTG CGCATCTCAC
8010 GTGTAGAGAT CACTTACCTA	8020 TTGGTGCACC AACACAGCTG	8030 GGTTTGTGAC CCAAACCTCG	8040 GGGGCGGCCA CCCCCGGGCT	8050 TACGTGACGT ATGCACTGCA	8060 CCCTTGCCCC GGGAACCGGG	8070 TGACCTTGTG ACTGGAACAA	8080 ACTGTCACTT TGACAGTGGG	8090 CTCGGGTCTT GAGCCCAAGG	8100 GAGCATGTGT CTCTGAACCA
8110 ACCTAGTAGT TGGATCATCA	8120 ACGAGCAGTA TGCTCGTCT	8130 CTATAGTTAC GATATCAATG	8140 AACCCTGTTC TTGGCAACAC	8150 TGTCCTGTGT ACAGGCACAC	8160 CAGCTATGTG GTGCATACAC	8170 AAGGAGTCTT TTCTCTAGGA	8180 AATGTTCGAG TTACAGCTTC	8190 GAGGGCGCAG CTCCCGCTGC	8200 TCTTGTGTATA AGAACCATAT
8210 GGGTCCCTTG CCCAGGGAAC	8220 TTGGGTAAGG AACCCATTCC	8230 ACTTAGTCGC TGAATCAGCG	8240 ATTTAGGCGT TAAATCCCAAC	8250 TGACGTCTCT ACTGCAGGGA	8260 TCTGGAGCGT AGACCTCGCA	8270 GCATTGAGTG CGTAACCTAC	8280 CAACACGTAA GTTGTGCATT	8290 CAGTTTCAAC GTCAAGGTGT	8300 ATGTAAGCCC TATACCTGGG
8310 GTGCTGCCCT CAGCAGCGGA	8320 ACTAGGAGGT TGATCTCCCA	8330 CATACCATCG GTATGTAGCG	8340 CGCCACAGAG GCGGTCTCT	8350 CAGAGTTTTC GTCTCAAAAG	8360 CTCCATCCCG GAGGTAGGCG	8370 TAGGCTATGAC ATCCCTACTG	8380 ATGCCCTACG TAGCGAGTGC	8390 CGGCTCTGTT GCGCAGACAA	8400 GGCTCTAGCA CGCAGCATGT
8410 CAACACGATC GTGTGTCGTA	8420 CAGAGTACGG GTGTCTATCC	8430 TTTACCTTGC AAATGGAACG	8440 GGCTGCGATC CCGACGTGTC	8450 AGTATAAAGG TCATATTTCC	8460 ACTTCTGTTT TGAAGCAAAA	8470 GGTCCACGCC CGGTGTCGCG	8480 CGCATGTGTT CGCTGACAAA	8490 GTCTAGACGC CAGATCTGCG	8500 AGAGGCCAGA TCTCCGGTCT
8510 GCAGCGAATC CGTGGCTTAG	8520 GAGCGAGACA CTCGCTCTGT	8530 CATCATCAAC GTAGTAGTTG	8540 ATCATATAGG TAGTATATCC	8550 TGAGAGAGTT ACTCTCTCAA	8560 TGTAGGTCTC AGCATCCAGG	8570 CGGGGGGACC CGCCCTCTGG	8580 GAACCCCAAG CTTCCGGTTC	8590 ATACATTAGA TATGTAAACT	8600 GGAAGTACCC CCTTCAATGG
8610 GGCGACGGGA CCGCTGCCCT	8620 CTATTGTAGG GATAACATCC	8630 TGTGGCGTTC ACCACCCGAG	8640 TTATTCGGTG AATAAGCCAC	8650 TGGGTCCGTT ACCCAGCCAA	8660 GGATGTGTAA CCTACACATT	8670 GCAAGACGCT CGTTCTCGGA	8680 CAGTGTGTGC GTCACACACG	8690 CCTCTCTGCC GGAGGAGCGG	8700 CTTCTCGACC GAAGAGCTGG
8710 TCTTGGTAC AAGAACATCG	8720 AAAAAATAAA TTTTTTTTTT	8730 AAATAGGTT TTTATTCCAA	8740 TCTAATAGG AAGATTATCC	8750 TTTGGAGTT AAAACTCCAA	8760 TTACTTCTAG AATGAAGATC	8770 ATAATTCTAT TATTAAGTGA	8780 TGCCGACGGG ACGCCCTCCC	8790 GAGGCCACCG CTCCGGTGGC	8800 CACCAGTTTC GTGGTCAAAC
8810 AGATGTCCGT TCTACGCCCA	8820 TCTTGTCTTA AAGAACAGAT	8830 TTACCGTAAA AATGGCATTT	8840 CATTTCTACA GTAAGATGTT	8850 CGTGTACCGG GCACATGGCC	8860 AAGGTTTTCG TTCCAAAAGG	8870 GTTTGACGGG CAAACTGCCC	8880 AGTGACGGTT TCACGTCCAA	8890 CACCCTGATT GTGGACGTAA	8900 TCCGATTTGG AGGCTAAACC
8910 GAACTCCAC CTTACGGGTG	8920 TTAGAGAGGA AATCTCTCT	8930 TATTTGTAAAG ATAAACATTTC	8940 GTCTGGAGAG CAGCACCTTC	8950 TTGGTACGGG AACCATGCCC	8960 TTTATTAATA AAATAATTTT	8970 GTAGAGCGGT CATCTGCCCA	8980 GGAATAGTTA CCTTATCAAT	8990 TACAGAGATT ATGTCTCTAA	9000 CGTTTAGGGC GCAATCCCGC

FIGURE 21 (Continued)

9010 TTATAATTC ATATAAAGT	9020 GGCCGGTAAC CCGGCCATTG	9030 ATTTTATGAC TAAATATCTG	9040 GAGGTCTCCG CTCCAGAGCG	9050 GGGAGGTGGA CCCTCCACCT	9060 AGTCGGAGTT TCAGCTCAAC	stop URF 24 start URF 23 9070 CGTCGGCTTAG GCAGCGAATC	9080 TACTAACGTT ATGATTGCAA	9090 TTTAAATCCA AAATTCAGGT	9100 AGGAGTGCTT TCCTCACAGA
9110 GGACATATTC CCTGTATAAG	9120 TAAGTTTTCG ATTCAAAAGC	9130 CCTGTAAATT GGAACATTAA	9140 GTTTATATGG CAAAAATACC	9150 CGCTAGGGCA CGCATCCCGT	9160 TCCAGGGGAA AGGTCCCTTC	9170 CGTCCCGGTC CGAGGGCCAG	9180 GACTTGATTT CTGAACATAA	9190 AGCACGTCCA TCGTGCAAGT	9200 GACGTGCGCTG CTGACGGAC
9210 GTCGCGCGCG GACGCGCGCG	9220 TGAAGGGGCG ACTTCCCGCG	9230 GTCTTGTGTA CAGGAACCAT	9240 CTGTTTCTCT GACAAAAGAA	9250 GGGTGTGACT CCCACTACTGA	9260 AATACGTGTC TTATGACAGC	9270 GTATGAGGCT CATACTGCGA	9280 CGATACGATT GCTATGCTAA	9290 GGTCCGATCG CCAGCTAGTC	9300 GGGATACATT CCTATGTAA
9310 CGAACAACTG GCTTGTGTGA	9320 ACCCGCGCGT TGGGCGGCGA	9330 ATATTTTACG TATAAATGCG	9340 TTCCAGCAGC AAGGTGCTGC	9350 AGTTTTTATG TCAAAAATTC	9360 TCCGTTTCGG AGGCAAGGCC	9370 AGCGCGTTTT TCGCGCAAAA	9380 TTCGTTCTGT AAGCAAGCAC	9390 TAGCATCAGT ATCGTAGTCA	9400 ACGAGTACGT TGCTCATGCA
9410 CTATTTCCGT GATAAAGGCA	9420 CCATTCAAGG GGTAAGTTCC	9430 CCTTGGTGGT GGAACCAACA	9440 GTCTTTTCTT CAGAAAAGAA	9450 GTGGTTAAAG CACCATTITT	9460 GAGAAGTTGT CTCTCAAAAC	9470 ACAGAGCGCC TGTCTGCGGG	9480 AAGGACGTAA TTCCTGCATT	9490 TTTGTGTTTT AAACACAAAA	9500 ATTTTATTGT TAAATAAACA
start URF 24 stop URF 25 9510 TTTTTTTTTT AAAAAAATAC	9520 TAAATTGTGA ATTTAACAT	9530 ATCTTCGGAC TAGAAGCCTG	9540 AGAAATGTTG TCTTCAACAA	9550 CCTTTTGTGT GGAAAACAAA	9560 GGGAATATTC CCCTTATAAG	9570 GTATTCGCGC CATAGACGG	9580 TGATGCCCGT ACTACGGCCA	9590 ACGGCCGCGC TGCCGGCGTG	9600 TGGCATTTTT ACGCTAAAAA
9610 TTGACCACTG AACTGGTCAC	9620 GCACATAATT CGTGTATAAA	9630 TTGCTGGTGG AAGCACCACC	9640 CTGTCAAGGA GACAGTTCTT	9650 GCCAGTACAG CGGTCATGTC	9660 GCCTCAGTAT CGAGTCTATA	9670 TACATTCTGA ATGTAAGACT	9680 GCCATTTGTG CGTAACACAC	9690 TAGTCCAAAC ATCAGGTGGG	9700 AATGTGAGCC TTAAGTATCG
9710 AGTCACGATT TCAGTGCTAA	9720 TTTGGCTGGC AAAGCGAGCG	9730 TTTATCGGGC AAATAGCCCG	9740 CCCCTTATGT GGGGAATACA	9750 ATGGGCGTCC TACCCGCAAG	9760 GCATCTCTGT CGTAGAGACA	9770 TGTAATGTGG ACATTACAGC	9780 GGGGTATCTT CCCATAGGAA	9790 CCATATGTTT GGTATAACAA	9800 TTAATTATCC AATTAATAGG
9810 TCTCTTTTTG AGAGAAAAC	9820 TGATTTTGTG ACATAAACAC	9830 GACTTTTTCG CTGAAAACCC	9840 GAGGACGGAT CTCTGCCCTA	9850 CCGTTTTATC GGCAAAATAG	9860 GTGGGAGGGC CACCCTCCCG	9870 GAGGTCTTGT CTCCAGAAAC	9880 TGATGTGCTG ACATACAGCG	9890 GAAGGTGTGG CTTCCACAGC	9900 CCGTGGTAT GGCAGCCATA
9910 TGTCAGTCCG ACAGTCACT	splice E4 leader 9920 AATGGTCATT TTACAGTAA	9930 TTTTTGGATA AAAAACCTAT	9940 ATTTTGTGTG TAAAAACAC	9950 GTGACCTGTG CACTCGACAC	9960 CCGTGGTCCA GGCAGCAGCT	9970 GTATGTCAGT CAATCAGTCA	9980 GTACATTTTT CAGTGTAATA	9990 TCCCGGTCCA AGGGCCAAAG	10000 TGCTCTGCTC ACAGAGCGAG
10010 ATATATATCC TATATATAGG	10020 TGATTTTTGA ACTAAAAAT	10030 CTGCAATGCC GACGTAAACG	10040 AATTTCAAGT TTAAAAGTCA	10050 GTTTGTGTGG CAAAAACAC	10060 GGTCTTTTGG CCAGAAAACC	10070 CGTGGCGCTG GCACGCGAAC	10080 GATGCGGCTC CTACGCCACG	10090 TTTGTCTTCG AAACGAAGCG	10100 GTTTTTGGGG CAAAAACCC
10110 TGTTGAAGSA ACACTTCTCT	10120 GTTTAGAAGT CAAACTTCCA	10130 GAAGGCAAAA CTTCCGTTTT	10140 GGGTGCTATG CCCACGATAC	10150 CAGTGAAGGG GTCACTTCCC	10160 TAAATTTTTC ATTTTAAAAA	10170 TTGATGTTAA AACTACAATT	10180 GGGTTATGTA CCCAATACAT	10190 CGTTCAATGA GCAAGTACT	10200 GGCGGGTTAA CCGCCCTAAA
10210 TGATGCAAGT ACCTACGCTA	10220 GGGCGGGGCA CCCGCCCGCT	10230 AGGGCGGGGG TCCACGCGCC	10240 GCGCGGCGCA CGCGCACAGT	10250 GTGTGTGAGG CACAACTCC	10260 TGGGGGAGTA ACCCCTCAT	10270 ATAGTATAGC TATCATATTG	10280 CGAAGTAGG GCTTCAATCC	10290 TTTATTCCCA AAAAATAGGT	10300 TATAAATACT ATATTATTGA
10310 ACTAC TGATG									

FIGURE 21 (Continued)

10 r3CCCCGGAAG 15GGGGACTTTC	20 TCCAACCAAT AGGTTGGTAA	30 CCACCCGCTC GGTGGGCGAG	40 AALCCATTAA TTGGGTAAAT	50 AAACAATTAA TTTGTAAATT	60 AGACAGAACG TCGTCTTTCG	70 TCGACGGTAC AGCTGCCAIG	80 TCACCTTCCG AGTGGAAAGC	90 GAAGAAAAC CTTCTTTTGA	100 CCCCCTCAT GGGGGGAGTA
110 AAATCGGGA TTTAGCCTT	120 TAGACTGCCC ATCTGACGG	130 GTCCGAGGCT CAGGTCTCCA	140 GTCACCGCTC CCATGGCGAG	150 CTCAAGCAGT GAGTTGCTCA	160 CTTACAGTAC GAATGTCTATG	170 CCTAGGTGAC GGATCCACTG	180 ACCTACCCCTC TGGATGGGAG	190 TGGGCGGTC ACCGCTCCAG	200 GGCGGGTTAA CCGCCCAATT
210 GGAGTTGCGA CCTCAACGCT	220 CTGGATACGG GACCTATGCC	230 TGAACCTCAA ACTTTGAGTT	240 GCAGTGGTAA CGTCACCAAT	250 CCTACGTGCA GGATGCAAGT	260 CGTCGGCGGC GCAGCCGCGC	270 GGCGATGAGC CCGCTACTGC	280 ACGGCGGTTG TCCGCGCAAC	290 TGGTAGAAC ACCATCTTTC	300 CTTACCCGAT GAATGGGCTA
310 AATACCTTCG TTATGAGAGC	320 TAGCAACGCT ATGCTTGCCA	330 TAGGTCAAG ATTCAGTTTC	340 GAGATTATGC CTCTAATAAC	350 GGAAGTTGGG CCTTCAACCC	360 ACGACTCTCT TGGCTGAGGA	370 GTTGATGAA CAAGCTACTT	380 CAAGAGAACC GTTCCTTGG	390 GGCTCGAGCT CGCAGCTGCA	400 CCGGAATTGG GGCCTTAACC
410 GTTGCGAATC CAACGCTTAG	420 CGCTTGACAG GGCACTGTCT	430 ATTCGCTCAC TAGCAGGTGG	440 CGGGTCAACG GCCAGTTTGC	450 CACTCGTATT GTAGCAAAAC	460 ACTGACACGA TGAGTCTGCT	470 CAACGGTGTG GTTGCCACAG	480 GTTTCAGATT CAAGTCTTAA	490 TTATTTCTAG ATAAAGATCT	500 GTTTAGTAT CAATCAATA
510 TTATTTCTTT ATAAAGAAA	520 ATGAACATA TACTTGATAT	530 TTTTGTGTAC AAACAATG	540 TTACAAAATA AATGTTTTAT	550 ACTAAAAGC TGATTTTTTC	560 GCG GCG				
			stop IVa ₂ protein						stop protein IX

FIGURE 22. Nucleotide sequence of a region between coordinates 9.6 and 11.2 on the Ad3 genome. This sequence was established by Engler (1981). The region codes for polypeptide IX. For the positioning of strategic signals, see Fig. 3 (Section VII).

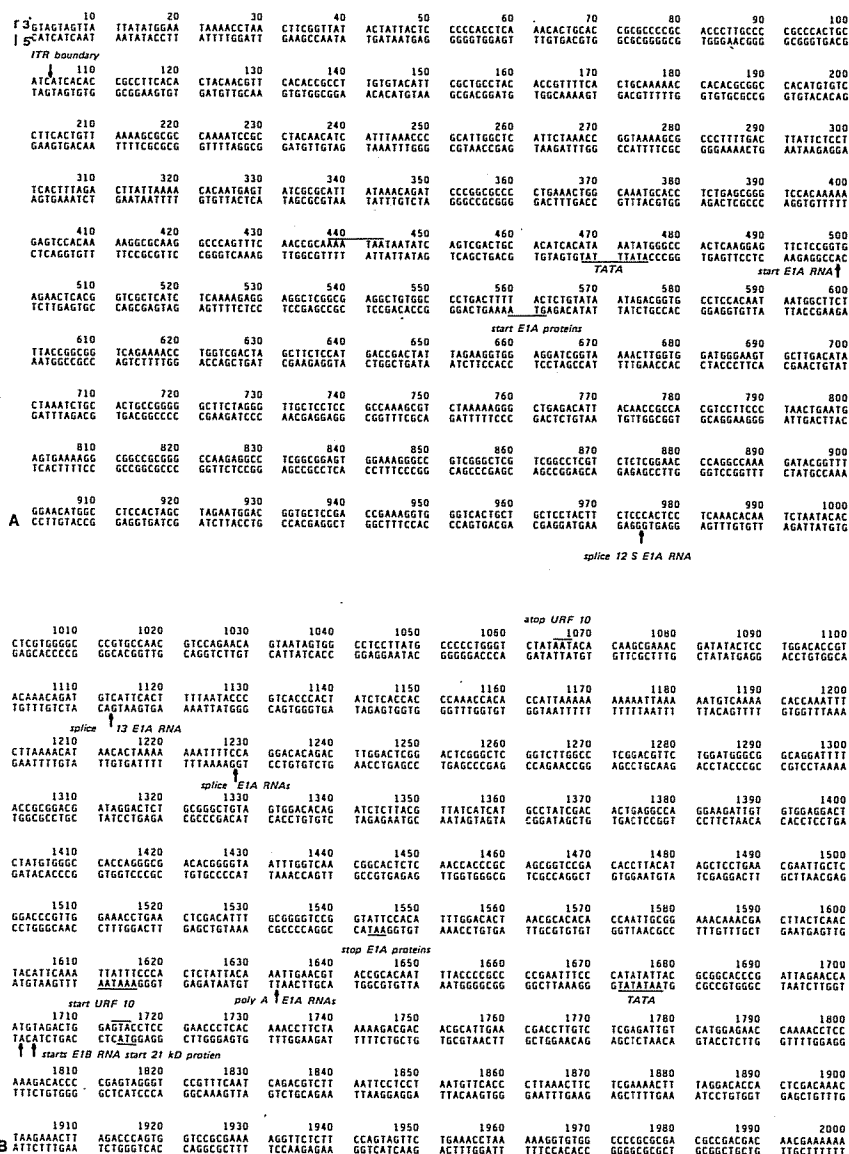


FIGURE 23.1A–L. Nucleotide sequence of a region between coordinates 0.0 and 31.7 on the Ad5 genome. This sequence was established by Steenbergh *et al.* (1977), van Ormondt *et al.* (1978), Maat and van Ormondt (1979), Maat *et al.* (1980), van Beveren *et al.* (1981), Bos *et al.* (1981), and H. van Ormondt and B. M. M. Dekker (personal communication). For interpretations, see van der Eb *et al.* (1979) and van Ormondt *et al.* (1980a,b).

2010 CTCAAAATAT GAGTTTATA	2020 TTCCTATTTA AAGGATAAAT	2030 CCTCGCTTCT GGAGCGAAGA	2040 TTGGGTAGAC AAACCATCTG	2050 TGGCCCECCA AGCGGGGGGT	2060 TGGACGACCT ACCTGCTGGA	2070 AAAGACCGGG TTTCTTGCCC	2080 TACGTAGACA ATGCATCTGT	2090 CCTCTCGCCA GGAGAGCGGT	2100 ACACTCTGTG TGTGAGACAC
<i>start 55 kD protein</i>									
2110 TCTTACCGG AGAATCGCC	2120 ACGATGCAA TGTACTGTT	2130 CAGAGCGAG GTCTTCCGT	2140 CGGGGCGCT CGCCCGCGGA	2150 ATTATGGCTG TAATACCGAC	2160 CCTCCTCGTC GGAGGAGCAG	2170 GTGCTGCTCG CAGCAGCAGC	2180 TCTCCTCTCG AGGAGGAAGC	2190 GTCCGCGGCC CAGGCGCGGG	2200 GCCGTCCTCG CGCGAGGAGC
2210 TCTCGGGTAC AGAGCCCATG	2220 CTTGGGCTCT GAACCGGAGA	2230 CGCCCGGACC GCCGCGCTGG	2240 TGGGAGCCCT ACCCCTGGGA	2250 TACTTACAAC ATGATGTTTG	2260 ATGTCACACG TACAGGTGGC	2270 ACTTGACATA TGAACGTGAT	2280 GGCTGTGACT CGAGAACTGA	2290 CTGGGTAAAA GACCACTTTT	2300 CTGTAAATGT GACAAATTACA
2310 CTCTACCGG GAGGTGCGG	2320 TCCCGGATTT AGGGGCTAAA	2330 CCCCCATTTT GGGGGCTAAA	2340 TCCCTGCGCC AGGGAGCGGG	2350 CGCGAACACT GGGCTTGTGA	2360 CGGATGCTCT GGCTACAGAG	2370 CTCCGATCCT GAGGCTAGGA	2380 TAGATCGAAA ATCTAGCTTT	2390 ATCGAATTAC TAGCTTAATG	2400 TGGTCTGTGG ACCGACACCC
<i>splice 13 S E1B RNA</i>									
2410 CAGGACTCAC GTCTGAGTG	2420 ATAATGAAA TATTACTTTT	2430 GTTGTCTAGT CAACAGATCA	2440 TCTATTAAAC AGGATAATTG	2450 CGGATTIACCT CGCTAATGAG	2460 GAACATAGAG CTGTATCTGC	2470 ACCGGCTCTT TGGCGCAGAA	2480 CATAGGATAT GTATTCCATA	2490 CTCGTGGACT GAGCAGCTGA	2500 GGTGAATGAC CCACTTACTG
2510 CGAGCTCGGT GCTGAGCCA	2520 CCCTACTATA GGGGATGATT	2530 AACTCCTCCG TTGAGGAGGC	2540 ATAATCCCAT TATTAGGGTA	2550 ATACGTTTCC TATGCAAAAG	2560 ACCGTGAATC TGGCACTTAG	2570 CGGTCCTAAC GCCAGATTGC	2580 TTCATGTTCT AAGTACAAGA	2590 AGTGTGTTGA TCAGCAAACT	2600 ACATTATATG TGTAAATATC
2610 TCTTAAACAA AGGAATGTT	2620 CGATGTAAG GCTACTTTC	2630 ACCTTTGCCG TGGGAACGGG	2640 CGGCTCCACC CCGAGGTGG	2650 TCTATGATAG AGATAGATAC	2660 CTCTCTATCC GGAGGATAGG	2670 CAGCGGAANT GTGGCTTTTA	2680 CTACATCGTA GATGTAGCAT	2690 CTATTATATC GATAAATATG	2700 ACCGCGCCCC TGGCGGGGGG
2710 TCCGACCGTA TGTGAGTTC	2720 CGTCCCGCAG GGACGGGGTG	2730 CAATAATACT GTTATTATGA	2740 TACATTCCAA ATGTAAAGTT	2750 ATGACCGGGG AATTTTAGCG	2760 TTAAATACGC AATTTTAGCG	2770 CATGCCAAAA GTACGGTTTT	2780 GGACCGGTTA CTGAGCCAAAT	2790 TGGTGTGAAAT ACCACTCTTA	2800 AGGATGTGAG TGTGACAGCG
2810 ACATTGGAAG TGTAAAGTTC	2820 ATACCCCAAT TGTATGGAC	2830 TGTATGGAC ACAATACCTG	2840 ACACCTTCCG TGTGGAAGCC	2850 ACCTGGCTAC TGGACCGATG	2860 ATTCCCAAGC TAAAGGTTCC	2870 CCGACACAGG GGGCTGTGCC	2880 AAATGACAGA TTTTACTGCT	2890 GAGCTTCCCG GCTGGAGGGG	2900 CCACCACACA GGTGTGTGT
2910 CGGGGGTTTT CGCCCCAAAA	2920 CGTCCCGAAG CGAGGGCTTC	2930 TAAATCTCTT AATTAAGAAA	2940 ACGGAGAAAC TGGCTCTTTG	2950 TTTCCACATG AAAGGTGTAC	2960 GAACCCATAG CTTGGGTATC	2970 GACAGACTCC CTGTCTGAGG	2980 CATTTAGGTC GTAACTCCAG	2990 CCACCGCGTG GCTGCCCAAC	3000 TTACACAGGA AATGTGGCCT
3010 GGCTGACACC CGCACTGGG	3020 AACGAAGTAC TTGCTTCATG	3030 GATCACTTTT CTAGTGAAAA	3040 CGCACCGACA CGGTGGCTGT	3050 CTAATTCGTA GATTAAGCAT	3060 TTGTACCATTA AACATGGTAT	3070 CAGCGTTGAC GTGGCACTGC	3080 GCTCTGTGCC CGAGGACAGG	3090 GGGAGGCTTC GCCTCTCAGA	3100 ACGACTGGAC TGTCTGACCT
3110 GAGCGTGGCG CTGGAGGGG	3120 TTGACAGTGG AAGTGTACCC	3130 ACGACTTCTG TGTGAAAGAC	3140 GTAAGTGCAAT CATTCACGTA	3150 CGGTGGGTGA GCCAGGCCACT	3160 GAGCGTTTCC CTGGCAAGGC	3170 GACCGGTTCG CTGGCCAGTG	3180 AACTCGGTAT TTTGAGCATA	3190 TGTATGACTG ACATACTGAC	3200 GGCGACAAAG CCGCTGTTC
3210 AACGTAACAC TTGCAATTGG	3220 CATTTGCTCT GTAACAGGAG	3230 CCCCCAAGAG GGGGGTGTTT	3240 GATGGAATGG CTACCTTACC	3250 TTACGTTAAA AATGCAATTT	3260 CTCAGTGTGA GAGTCACACT	3270 TTCATATACG AAGATATTGC	3280 AACTCGGGCT TTGAGCCGGA	3290 CTGCTACAGG GAGCATGTCC	3300 TTCCACTTGG AAGGTGAACC
3310 ACTTGCGCCA TGAGAGCGGT	3320 CAAACTGTAC GTTTGACATG	3330 TGGTACTTCT ACCATGAAGA	3340 AGACCTTCCA TCTGGAAAGT	3350 CGACTCCATG GCTGAGGTAC	3360 CTACTCTGGG GATGAGACCC	3370 CGTGGTCCAC GCACCAGGTG	3380 GTCTGGGACG CAGACCTTGC	3390 CTCACACCCG GAGTGTGGGG	3400 CATTTGTATA GTAAACATAT
3410 ATCCTTGGTC TAGGAACCAAG	3420 GGACACTACG CCTGTGATGC	3430 ACCTACACTG TGTATGTGAC	3440 GCTCCTCGAC CGAGGAGCTG	3450 TCCGGGCTAG AGGCCCGATC	3460 TGAACACAGA ACTTGGTGCT	3470 CCGGAGCTGG GGCCTGCACC	3480 GGCGGACTCA CGGCTGAGT	3490 AACCAGATGC TTGGCTCTAG	3500 GCTACTCTTA CGATGAAGAT
3510 TGTCTAACTC ACAGATTGAG	3520 CATGACTTTA GTACTGAAAT	3530 CACACCCGCA GTGTGGCGGT	3540 CCGAATTECC GGCTTAAAGG	3550 ACCTTTTCTT TGGGAAGAAA	3560 ATATATTCCA TATATAAGGT	3570 CCCCCAGGAT GGGGCTCTTA	3580 ACATCAAAAC TGTAGTTTTT	3590 ATAGACAAAA TATCTGTTTT	3600 CTGCTGCGGC GGAGCAGCCG
3610 GGGGGCGGTA CGGCGGCTAT	3620 CTCGGTGTTG GAGCACCAC	3630 AGCAAGACTAC TCGTTTGATG	3640 CTTCGTAAAC GAAGCAATGT	3650 CTCGAGTATA GAGTCATAT	3660 AACGTGTTGG TTGACAAACG	3670 CGTACGGGGG GCATGCCCCC	3680 TACCGGCCCC ATGGGCGGGG	3690 CAGCGAGTCT GTGCTCAGAA	3700 TACACTACCC ATGTGATGGG
<i>splice 22 S E1B RNA</i>									
<i>start protein IX</i>									
3710 GAGGTGCTAA CTCCAGCAAT	3720 CTACCAAGCG GATGGTGGCC	3730 GGCAGGACGG CGTCTCTGCC	3740 GGCTTTGAGA CGCAAACTCT	3750 TCAATGAACT ACTACCTTGA	3760 GGATGCTGCT CCTACGAGAC	3770 GCACAGACCT CGTGTCTGGA	3780 TGGCGCAACC ACCGCGTTGG	3790 TCTGAGGTGG AGACTGAGCG	3800 GAGCGGCGGG CTCCGCGGCC
3810 CGAAGTCGGC GCTTACCGCG	3820 GACGTCGGTG CTGAGGCCAC	3830 CGCGGCGGCC CGCCCGCGGG	3840 TAACACTGAC ATTGTGACTG	3850 TGAACGAAA ACTTTGCTTT	3860 GGACTCGGGC CCTGAGCGCG	3870 GAACGTTGCT CTTGCAAGCA	3880 CACGTCGAAG GTGCGACTTC	3890 GGCAAGTAGG CCGCTGATCC	3900 CGGCGGCTAC GCCCGCGATG
3910 TGTTCACACT ACAAGTTTGC	3920 CCGAATGAAC GGCTCTTTTG	3930 CGGTGTAACC GCACAATTGG	3940 TAAAGAACTG ATTCTTTGAC	3950 GGCCCTTGAA CCGGGAACCT	3960 TTACAGCAAA AATGTGCTTT	3970 GAGTCGTGGA CTCAGCAACT	3980 CAACCTAGAC GTTGGATCTG	3990 GCGGTGCTCC CGCCAGCAGG	4000 AAAGACGGGA TTTCTGCCCT

FIGURE 23.1 (Continued)

4010	4020	4030	4040	4050	<i>polyA protein IV₂ RNA</i>		4080	<i>stop protein IV₂</i>	
CTTCCGAGG	AGGGGAGGT	TACGCCAAT	TTTGTATTA	TTTTTTGGT	TGAGACAAAC	CTAAACCTAG	TTCGTACAC	GAACGACAG	ATAAATCCC
GAAGGCTTC	TECCCTCCCA	ATCGGTTTA	AAACATAAT	AAAAAACCA	ACTCTGTTG	GATTGGATC	AAGCAGGTGT	CTTGTGTCT	TTATTAGGG
		<i>stop protein IX</i>							
4110	4120	4130	4140	4150	4160	4170	4180	4190	4200
CAAAACGCG	GGCCATCCG	GGCCCTGGT	GGCAGAGCCA	GCAACTCCCA	GGACACATA	AAAAGCTCT	GCACATTTC	CACTGAGACC	TACAAGTCTA
GTTTTCGGC	CGCGGTAGG	CGCGGACCA	CGGTCTCGT	CGTTGAGGG	CCTGTGTATT	TTTTCCAGG	CGTGGTAAAG	GTGACTCTGG	ATGTTCCAGT
						<i>polyA⁺ E1B RNA₂</i>			
4210	4220	4230	4240	4250	4260	4270	4280	4290	4300
TGTACCGTA	TTGGGACGA	GACCCACCT	CCATCTGCT	GACGTCTCA	AGTAGCGCC	CCCACACAA	CATCTACTAG	GTACGACATC	TCTCCGACAC
ACATGGGCA	AAAGCCGTCT	CTGGGTGGA	GGTAGACCA	CTGCAGAGCT	TCATGCTCG	GGGTGGTGT	GTAGATGATC	CAGTCTAGC	AGGAGCCGTG
4310	4320	4330	4340	4350	4360	4370	4380	4390	4400
CGGCACGAC	GATTTTTTACA	GAAGTCATC	GTTCGACTAA	CGGTCCCGGT	CGCGGAACCA	CATTACACAA	TGTTTTGCCA	ATTGACCCCT	ACCCACGTAT
GGCGTGTGG	CTAAAAATGT	CTTTCAGTAG	CAAGCTGATT	GCCAGGGGCA	GGCCCTTGGT	GTAACTGTTT	ACAAAGCGGT	TAAGCTGGGA	TGGTGCATA
4410	4420	4430	4440	4450	4460	4470	4480	4490	4500
GCACCCCTAT	ACTCTACGTA	GAACCTGACA	TAAAAATCCA	ACCGATACAA	GGTCCGTAT	AGGAGGCGCC	CTAAGTACAA	CACGCTTGG	TGCTGTGTC
CGTGGGATA	TGAGATGCA	CTTGGACTGT	ATTTTATAGT	TGGCTATGTT	CCGAGCCATA	TCCCTCCGGG	GATTCTGTT	GTGCAAGACC	ACCAAGACAG
4510	4520	4530	4540	4550	4560	4570	4580	4590	4600
ACATAGGCA	CGTGAACCT	TAAACAGTA	CATCGAATCT	TCCTTTACGC	ACCTTCTTGA	ACCTCTGCGC	GAACACTGGA	GGTCTAAAA	GGTACGTAA
TGTATCCGT	GCATTTGGGA	AATTTGTGAT	GTAGCTTAGA	AGGAAATGCG	TGGAAGAACT	TGGAGAGCCG	CTTGTGACCT	CCAAAGATTT	CTTACGATTC
4610	4620	4630	4640	4650	4660	4670	4680	4690	4700
CAGGTATTAC	TACCTTAAAC	CGGGTGAAGC	CGCCGCGGCG	CGCTTCTATG	AAGAGCTGCT	TGATTGCTAC	ATCAACACAA	GGTCTTACTC	TAGCAGATC
GTCTCAATG	ATGGCAATGG	GCCACAGGCG	GGCGGCTTGG	GGCAAGATAT	TTCTGGGATC	ACTAACGTCA	TAGTTGTGTT	CCAGGATGAG	ATCTCATAG
4710	4720	4730	4740	4750	4760	4770	4780	4790	4800
CGGTAATAAT	GTTTCGCGCC	CGCCTCCAC	GGTCTGACGC	CATATTACCA	AGGTAGGCGC	GGTCCCGGCA	TCAATGGGAG	TGTCTAAACG	TAAAGGGTGC
GGCATTTTTA	CAAGCGCGGG	CGCGAGGGTG	CCAGACTGCG	GTATAATGGT	TCATCCCGGC	CCAGGGGCGT	AGTACCCTC	ACAGATTGCG	ATTTCACAG
4810	4820	4830	4840	4850	4860	4870	4880	4890	4900
GAACCTCAG	TCTACCCGCC	TAGTACAGAT	GGACGCCGCG	CTACTCTCTT	TGCCAAGGCG	CCCATCCCTT	CTAGTCGACC	CTTCTTTGCT	CCAAGGTGAC
CTTTAGGTC	AGATGGGGGG	ATCATGTCTA	CTCTGGGGCG	GATGAAGAAA	ACGGTTTCCG	GGGTAGGGGA	GATCAGCTGG	GAAGAAGACA	GGTCTCTGAC
4910	4920	4930	4940	4950	4960	4970	4980	4990	5000
GTGACGCTG	AATGGCGTGG	GGCACCAGGG	CATTTAGTGT	GGATAATGGC	CCACGTTGAC	CATCAATTTT	CTCGACGTGG	ACCGGAGTAG	GGACTCGTGC
CAGCTCGAC	TACCGGACG	CGGTGGGGCG	GTAAATCACA	CTTATTACCG	GGTCAACTG	GTAGTAAAGA	GAGTGGAGC	TGGCTGCATC	CTTGAGGAGG
5010	5020	5030	5040	5050	5060	5070	5080	5090	5100
CCCCGGTGAA	GCAATTCGTA	CAGGAGCTGA	GGGTACAAAA	GGGACTGGTT	TAGGCGGTCT	TCCCGGAGCG	GGGGGTGCT	ATGCTCAAGA	ACGTTCTTTC
GGGGCCACTT	GTATTAAGAT	GTCCCTGACT	CGCATGTTTT	CCCTGACCAA	ATCCGCCAGA	AGGCGCTCCG	CGCCACGCGA	TAGCAGTTCT	TGCAAGGAAG
5110	5120	5130	5140	5150	5160	5170	5180	5190	5200
GTTTCAAAAA	GTTCCTAAAC	TCTGGGAGCG	GGCATCGCTA	CGAAACTCG	CAAACTGGTT	CGTCAAGGTC	CGCCAGGGTG	TGAGCCCAAT	GGACGAGATG
CAAAAGTTTT	CAACGGTTTT	AGACCGTCCG	CGGTATGAGC	GGTTTGGAGC	GTTTGACCAA	GCAGTTCCAG	CGCGTCCAC	AGCTCGGTCA	CTGTCTTAC
5210	5220	5230	5240	5250	5260	5270	5280	5290	5300
CCGTAGAGCT	AGGTGCTATA	GAGGAGCAAA	GCACCCCAAC	CGCCCGAAG	CGCATGCGCG	TCATCAGCCA	CGAGCAGGTC	TGCCCGGTTC	CAGTACAGAA
GGCATCTGGA	TCCAGCATAT	CTCCTGCTTT	CGCGGGTTGG	GGCGGCTTTC	CGTGTACGGC	AGTAGTGGGT	GTCTGTCCAG	ACGGGCGAGG	GTCTGTCTTT
5310	5320	5330	5340	5350	5360	5370	5380	5390	5400
AGGTGCGCGC	TGCCCAGGAG	CAGTGCATC	AGACCCAGTG	CCACTTCCCG	ACGCGAGGCG	CGCGGCGGCT	CGCGTCCGAC	GGCACTCTCG	ACCAGAGACA
TCCAGGGGCG	CAGGGTCTCT	GTGAGCGTAG	TTGGGGTCA	GGTGAAGGGG	TGCGCTCCGG	CGTCCGCGCT	GGCCAGGGTG	CGGTGTGAGC	TGGTCTCTCT
5410	5420	5430	5440	5450	5460	5470	5480	5490	5500
CCACGACTTC	GGGACGGCCA	GAGCGGGGAC	CGCGACCGCG	TCCATCGTAA	ACTGGTACCA	CAGTATCAGG	TCCGGGAGGC	CGCCACCGCG	GAACCGCGCG
GGTGTGAAG	CGGTGCGCGT	CTTCGCGCTG	CGCGTCCGCG	AGGTAGCATT	TGACCATGGT	GTCAATGTC	AGCCCTCCG	CGCGTGGCG	CTTGGGCGCG
5510	5520	5530	5540	5550	5560	5570	5580	5590	5600
TGGAACGGGA	ACCTCTCTCG	CGCGGTGCTC	CCCGTCAAGT	CTGAAACTCT	CGCATCTCG	AACCCGCGCT	CTTTATGGCT	AAGGCCCTCT	ATCCGTAGGC
AGCTTGCCCT	TGGAGGAGGC	GGCGACAGAG	GGCGAGTGCA	GACTTTTGA	GGCGTAGAGC	TTGGGCGCGA	GAATACCGA	TTCCGGGGAG	TAGGCATCGC
5610	5620	5630	5640	5650	5660	5670	5680	5690	5700
CGCGCGTCCG	GGCGGTCTGC	CAGAGCGTAA	GGTGTCTGGT	CCACTCGAGA	CGCGCAAGCC	CCAGTTTTTG	GTCCAAAGGG	GGTACGAAA	ACTACGAAA
CGCCGAGGCG	CCCGCAGACG	GTCTCGCATT	CCAGGAGCCA	GGTGAAGTCT	GGCGGTTCCG	GGTCAAAAC	CAGGTTTCCC	CCATGCTTTT	TGATGCTTTT
5710	5720	5730	5740	5750	5760	5770	5780	5790	5800
GAATGAGAC	CAAAAGTACT	CGGCCACAGG	TCCGACGAC	TGCTTTTCCG	ACAGGACAG	GGGCATATGT	CTGAATCTCT	CGGACAGGAG	CTGCCACAAA
CTTACTCTGT	GTCTCATGTA	CGCGGTGTCC	ACGCTCGGTG	ACGAAAAGGC	TGCTCTGTCT	CCCGTATACA	GACTTGAGAG	GGCTGTCTCT	GAGGGGTGTT
5810	5820	5830	5840	5850	5860	5870	5880	5890	5900
GGCGCAGGA	GGAGCATATC	TTTGAGCCTG	GTGAGACTCT	GTTCCTGAGC	GCAGGTCCGG	TCGTGCTTCC	TCCGATTCAC	CTCCCTCATC	GGCAGCAACA
CGCGGTCTCT	CTCTGTATAG	AAACTCGGAC	CACTCTGAGA	CAAGGCTCG	GTCCAGGCGC	AGCCAGAGG	AGGCTAAGTG	GGAGGGTAG	CGGTCTGTGT
5910	5920	5930	5940	5950	5960	5970	5980	5990	6000
GGTGATCCCC	CAGGTGAGCG	AGGTCCGACA	CTTCTGTGTA	CAGCGGGAGA	AGCCGTAGTT	CTTCCACTCA	ACCAACATC	CACATCCGGT	GCATGGCCC
CCACTAGGGG	GTCCACTCGC	TCCAGGTGTT	GAAGACACAT	GTCCCTCT	TCGGCATCAA	GGAAAGTGAT	TGGTTGTATG	GTGTAGGCCA	CGTAGCCGGG

FIGURE 23.1 (Continued)

6010 ACAAGGACTT TGTTCTGAA	6020 CCCCCGGATA GGGGGGGCTAT	6030 TTTTCCCGCA AAAAGGGGGT	6040 CCCCCGGCGA GGGGGGGCGT	6050 AGCAGGAGTG TCGTCTCTAC	6060 AGAGAAAGGG TCCTCTCCGC	6070 TAGCGACAGA ATCGCTGTCT	6080 CGCTCCCGGT GCGAGGGCCA	6090 CGACAAACCC GCTGTGGGG	6100 ACTCATGAGG TGAGTACTCC
6110 GAGACTTTTC CTCTGAAAG	6120 GCCCCGCTAG CGGGCATGAC	6130 AAGACGCGAT TTCTGCGCTA	6140 TCTAACAGTC AGATTGTGAC	6150 AAAGGTTTST TTTCCAAAAA	6160 GCTCTCTCTA CGAGGAGGAT	6170 AACTATAAGT TTGATATTCA	6180 GGACCGGGCG CCTGGCCGCG	6190 CCACTAGCGA GGTATGGCT	6200 AACTCCACCC TTGAGGGTGG
6210 GGCGTAGGTA CCGCATCCAT	6220 GACCACTCTT CTGGTCAGAA	6230 TTCTGTTAGA AAGACAATCT	6240 AAAACACACG TTTTGTGTCT	6250 TTCGAACCCG AAGCTTGGTG	6260 CGTTTGTCTG GCAACGACCC	6270 GCATCTCCCG CGTAGAGGCG	6280 CAACCTGTCT GTGGACAGCG	6290 TTGAAACCGT AACTTGGCGA	6300 ACCTCGCGTC CTGAGCGCAG
6310 CCAACCAAAA GGTTTGGTTT	6320 AACACGCGCTA TTGTGCGAT	6330 GCCCGCGGAG CGGGCGCTCT	6340 GAACCGGCGC CTTGGCGCGC	6350 TACAAATCGA ATGTTTAGCT	6360 CGTGCTTAAG GCACGTATTCT	6370 CGCGCGTTTC GGGCGCAACG	6380 GTGGCGGTAA CACCCGCATT	6390 GCCCTTCTCT CGGGAAGAC	6400 CCACACACCG GGTGTGCGC
6410 AGCAGCGCGT TCGTCCGGCA	6420 GGTCCACGTG CCAGGTGAC	6430 CGCGGTTGGC GCGCAACCCG	6440 GCCAACACGT CGGTTGTGCA	6450 CCCACGTGTC GGGTGCAAG	6460 CAGTTGCGAC GTCAACGCTG	6470 CACCGATGGA GTGGCTACTT	6480 GAGGCGCATC CTCCGCTAG	6490 CGGAGCAAC CGCTCTGTT	6500 CAGGTGCTCT GTCAACGAGA
6510 CGCCCGGCGG GCGCGCGCGC	6520 GAACCGGCTC CTTGGCGGAG	6530 GTCTATACCG CAGAAATGGC	6540 CATCCCGCAG GTAGGGGGTC	6550 ATCGACGACG TAGCTGCGTC	6560 AGCAGGCGCC TCGTCCGGGG	6570 CCAGACGCGG GGTCTGCGTC	6580 GTGGCGGTAA CAGCGTAAAG	6590 TGGGGCGCGT ACCCCGGGCA	6600 CGTCCGCGCG GCAGGCGCGC
6610 CAGCTTCTAT GTGGAAGTAG	6620 AGATAGAAGC TCTATTCTGC	6630 TAGGAACGTT ATCCCTTGAA	6640 CAGATCGCGG GTCTAGCGCC	6650 ACGACGGTAT GTCTGCGCAT	6660 GCGCGCGCGC GCGGGCGCGC	6670 TTGCGCGCGC AAGCGCGCGC	6680 AGCATACCCA TCGTATGGCT	6690 ACTCACCCCG TGAGTGGGGG	6700 TGGGCTACCG ACCCATATGG
6710 TACCCACCCC ATGGGTTGGG	6720 ACTCGGCGCT TGAGCGCGGA	6730 CCGATGTATC GGCGTAGATG	6740 GGCGTTTACA CCGCAATGCT	6750 GCATTGTGCA CGTAAACGTA	6760 CTCCCGGAGA GAGGGGCTCT	6770 GACTCATTAAG CTGATATTCT	6780 GTTCTATACA CAAGATATGT	6790 TCCCATCGTA AGGGTAGCAT	6800 GAAGGTGGCG CTTCCACCGC
6810 CCTACGACCG GGATGCTGGC	6820 CGCGTGCAAT GCGCACGTAA	6830 AGCATATCAA TCGTATAGTT	6840 GCACGCTCCC CGTGGGAGGG	6850 TCGCTCTCTC AGCGAGGAGG	6860 AGCCCTGGCT TCGGGACCGA	6870 CCAACGATGC GGTTGCTAGC	6880 CGCCCGGACG GGCGGGCTGC	6890 ACACGACGCT TCTGCTGGA	6900 TCTGATGAC AGACTATCTG
6910 GGACTTCTAC GCTTGAGATG	6920 CGTATACCTA GCATGTGAGT	6930 ACCTATATAT TGGATGATAT	6940 CCAACCTGCG GGTTGGAGCG	6950 ACCTTCTGCA TGGAAAGCGT	6960 ACTTCGACCG TGAAGTGGC	6970 CAGACACTCT GTCTGTGAGA	6980 GGATGGCGCA CCTACCGCGT	6990 GTGCGTGGTT CACGACGAGA	7000 CCTCCGCTAT GGAGGCGTAG
7010 CTACGCGCGT GAGTGGCGCA	7020 CGAACAAGTG GCTTGTGAC	7030 GTGAGCGCGC CAGCTGCGGC	7040 CACTGGAGCT GTACCTGCTA	7050 GCAGATCCCG CGTCTAGGCG	7060 CGTCAATCAGG GCAGTAGTCC	7070 TCCCAAGGGA AGGGTTTCTT	7080 ACTACTACAG TGATGATGTC	7090 TATGATATGG ATACTATACC	7100 ACAGCGGAAA TGTCTCTTTT
7110 AAAAGGTGTC TTTTCCACAG	7120 GAGCGCCCAAC CTCGCGGTTG	7130 TCTGTGTTGA AGGACAACCT	7140 GAAGCGCGCAG CTTCCGCGTC	7150 AAAGGTGCTG TTTCCAGTAC	7160 AGAAGCTAGC TCTTGGATCG	7170 CTTTGGGCGG GAAACCGCTC	7180 CGGAGGGCTT GGCTCCCGAA	7190 GCCATCTCTG CGGTAAGAGC	7200 GATCGTACAT CTAGCATGTA
7210 CTTGACCAAC GAATGCTGTT	7220 TGCCGGGACCA ACGCGCTGGT	7230 TCCGCGTCTG AGGCGACGCA	7240 AGGGAAAAAG TCCCTTTTCT	7250 TGCCCATCGC ACGGGTAGCG	7260 GCATACGGAC CGTATGCTCT	7270 GCGCGGGAA CGCGCGCTTC	7280 GCCTCGCTCC CGGAGCGAGG	7290 ACACCCATCT TGTGGTGA	7300 GCGTTTCCAC CGCAAGGTTG
7310 AGGGAAGTGT TCCTGACCA	7320 ACTGAAGCTC TGACTTTGAG	7330 CATGACCTATA GTACTGGTAT	7340 AACTTCACTC TTGAAGTCAG	7350 ACAGAGAGCT TGTGCTGCGA	7360 AGCGGGGAGC TCCGCGCTGC	7370 AGGGTCTCTG TCCAGAGCA	7380 TTTTGAGGCA AAAAGTCCGT	7390 CGCGAAAAAC GCGCTTTTTG	7400 CTTGGCGCTA GAACGCGGAT
7410 AAGCGTCCCG TTGGCAGGGC	7420 CTTCCACTGT GAAGGTGACA	7430 AGCAACTTCT TCGTTGAAGA	7440 CATAGAAAGG GTATCTTTCC	7450 GCGCGCTCCG CGCGGAGGGC	7460 TATTTCAACG ATAAGTTGCG	7470 CACACTACGC GTGTGATGCG	7480 CTTCCAGGGC GAAGGTCTCC	7490 CGGTGGAGCC GGCACTGCGG	7500 TTGCCAACAA AACGGTTGTT
7510 TTAATGAGCC AATTACTCGG	7520 GCGCGCTGCT GCGCGGAGCA	7530 GCTAGAGCAG CGATCTGCTC	7540 TTTGGGCAAC AAAGCCGTTG	7550 TACAACACCG ATGTTGTGGC	7560 GGTGTATACAT CCACAATGTA	7570 TTCAAGGTTT AAGTTCCAG	7580 TTCGCGCCCT AAGCGCGGGA	7590 ACGGGAACCTA TGCCCTTGAT	7600 CCTTCCGTTA GGAAGGCAAT
7610 AAAAATTCAA TTTTTAAGTT	7620 GGAGCATCCA CCTGTAGGTT	7630 CTCGAGAAGT GAGCTTCTCA	7640 CCCTCTGACT GGGGAGCTGA	7650 CGGGCAGCAG GCCGTGCTCT	7660 ACTTTCCCGG TGAAGGGGCG	7670 GTGAGACGTT CAGCTCTGCA	7680 CTACTCCCAA GATGAGGGTT	7690 CCTTCCGCTG GGAAGCGAGC	7700 TTACTCGAGG AATGAGCTCC
7710 TGTCGAGTGC ACAGGCTACG	7720 CCGGTAATCG GGCATTAGC	7730 TAACGCTCCA ATTTCAGGTT	7740 CCAGCGCTTT GGTCCGCAAA	7750 CCAGGATTTG GGTCCCTAAC	7760 ACCGCTGGAT TGGCGACCTA	7770 ACCGGTAATA TGCGCATTTT	7780 AAGACCCAC TTCGCGGGTG	7790 TACGTCATCT ATGACGTAGA	7800 TCCATTGCGC AGGTAAGCGG
7810 CAGAACAAAG GCTTCTTCC	7820 GTCCGCAAGG CAGCGGTCCC	7830 TAGGTTCCAA ATCCAAAGTT	7840 GCGCCGATCC CGCGGCTAGG	7850 AGAGCGCGCC TCTCCGCGCG	7860 GTCAATGATC CAGTCACTAG	7870 TCCGAGTAGA AGGCTCATCT	7880 GCGCGCTTGA CCGCGCAACT	7890 AGTACTGGTC TCATGACCA	7900 GTACTTCCCG CATGAAGGCG
7910 TGCTCGAGCA HACAGGCTGCT	7920 AGGGTTTCCG TCCCAAGGCG	7930 GGGGTAGGTT CCCCATCCAA	7940 CATATCCAGA GTATAGGCTCT	7950 GATGTAGCAT CTACATCGTA	7960 CCACTGTTTC GGTGACAAAG	7970 TCTCGGAGCC AGACGCTCGG	7980 ACGCTCTTAC TGGCAGGATG	7990 GCTCGGCTAG CGAGCGGATC	8000 CCCTTCTTGA GGGAAGCACT

FIGURE 23.1 (Continued)

8010 CCTAGAGGCG GGATTCCTCG	8020 GGTGGTTAAC CCACCAATTG	8030 CCTCTACCGC GAGGAGTGCG	8040 ATACTACAC TATTGATGTG	8050 CACTTTTCATC GTGAAAGTAG	8060 TTCAGGGACG AAGTCCCTCG	8070 CTGCCCCGGT GACGGGCGGA	8080 TGTGAGCAGC ACACTCTGTC	8090 ACCGAAACA TGGCTTTTGT	8100 TTTTTGACG AAAAAGTGC
8110 GCTCATACC GCAGTACTGG	8120 GTCCGCCAGT CAGCGGTGCA	8130 GCCCAGCATG CGGCGTGAC	8140 TAGGACGTGC ATCTCTGACG	8150 TCCAAGTGGA AGGTTGACCT	8160 CTGCTGGCGC GAGGACCGCC	8170 GTGTCTCTTC CACAAGGAGG	8180 GTCTACCTCT CAGAGTGGA	8190 TAAACTCGGC ATTGAGCC	8200 GACCGGACCG CTCGCTGCG
8210 CCCAAGCGA GGGTTTGGCT	8220 CCACAGGAAG GGTGGTCTTC	8230 ATGAAGCCGA TACTTCGGCT	8240 CGAAGAGAA GCTTGTCTTT	8250 CTGGCAGACC GACCGTCTGG	8260 GACGAGCTCC CTGCTCGAGG	8270 CCTCAATGCC GGAGTTACGG	8280 ACCTAGCCTG TGGATCGGAC	8290 GTGGTGGGCG CACCACGCGC	8300 GCGCTCGGGT CGCGAGCCCA
8310 TTCAGGGTCTA AAGTCCAGAT	8320 CAGGCGGCGG GTCCGCGCGC	8330 CCGCCAGGCT GGCGGTGGA	8340 CGAATCTACTG GCTTGATGAC	8350 TGTAGCGCGG AAATCTCGCG	8360 TCTACCTCTG AGATGGGAGC	<i>1st ATG 120 kD proteins</i>			
8410 CTGAGGAGCG GAGCTCTGCG	8420 TCCAAATGGA AGGTTTACCT	8430 GGGTATCTGC CGCATAGACG	8440 CCAGTCCCGC GCTCAGGGCG	8450 GCCCGATCTA CGGGCTAGAT	8460 GGTCCACTAT CCAGGTGATA	8470 GGATTAAAGG CCTAATTTC	8480 TCCCGGACCA AGGGGCTGGT	8490 ACCACGCGCG TGGTGGCGCG	8500 CAGTACCGA GTCTAGGCT
8510 ACGTTCTCTC TGCAAGAGGC	8520 CGGTAGGGGCG CGCATCTCCG	8530 GCCGCGCTGA CGGCGTACCG	8540 TGCCATGGCG ACGGTACCGC	8550 GCCGCGCGCG GCGCGCGGCG	8560 CACCGCGCGC GTGGCGCGCG	8570 CCCCAGGAGG GGGGTGTCTT	8580 ACCTAGTACG TGGATGATGC	8590 TAGATTCTCG ATCTAAAGC	8600 CCAGTCTCGC GGTACGCGCG
8610 CGCTCGGGCG GGGAGCGCCC	8620 CCCTCATCTCC GGGAGTACCG	8630 CCCCGAGGCG ACCCGCGGCG	8640 TGGGCGGCGC ACCCGCGGCG	8650 TCTCCGCGCG AGAGGGGCGA	8660 CCCCGTTGAG GGGCGCACGC	8670 CCCGGCGCGC GGGCGCGGCG	8680 CGCCGCTGAG GCGGCGGAGG	8690 CGACACGACG GCTGGTGTGT	8700 GCGCGCATCC GGGCGGTAGG
8710 AACGACCGCT TTCGTGCGGA	8720 TGGCGTCTGC ACGGGAGGAC	8730 CGCGCGCCAG CGCGCGGTGT	8740 TAGAGGACTT ATCTCTGAA	8750 AGACCGCGGA TCTGCGCGCT	8760 GACGCACTTC CTGCGTGAAG	8770 TGTGCGCGCG ACGACGGGCG	8780 GCCACTCGAA CGGTGAGCTT	8790 CTCGGACTTT GAGCGTGAAA	8800 CTCTCAAGCT GAGAGTTGCA
8810 GTCTTAGTGA CAGATCAAT	8820 AAGCCACAGC TTCGTGTCTG	8830 AAGTCCGCGC TTCAGCGGCG	8840 GGACCGCGCT CTGCGCGA	8850 TTAGAGGAGC AATCTCTGCG	8860 TGCAAGGAGC ACGCTCTCTG	8870 TCAACAGAAC AGTTGTCTTG	8880 TATCCGCTAG ATAGGCGATC	8890 AGCCGCTACT TGCGCGATGA	8900 TGACGAGCTA ACTGCTGAT
8910 GAGAGGAGCG CTCTTCTCTC	8920 ACCTCTAGAG TGGAGATCTC	8930 GCGCAGGCGC CGCGTCCGCG	8940 AGCGAGGTGC TCCGCTCCAG	8950 CACCCGCGCT GTGGCGCGGA	8960 CCAGCACTCT GGTCTGTGGA	8970 TTACGCGCGC AATGCGGGCG	8980 TACTCGAGCG ATGAGCTGCG	8990 TCTTCCGCAA AGAAAGCGGT	9000 CTCCGGAGGG GAGGCGTCCC
9010 AGCAGAGGCT TCGTTCGAGA	9020 GCGCCGACAT CGCGGCTGTA	9030 CTGGTGGCGG GACCAAGCCG	9040 GGAAAGCGTA CTTTGCGCAT	9050 GCGCCCGCGC CGCGGGCGCG	9060 GTACTGGTGG CATGACCACC	9070 ACGCGCTCTA TGCGCGGAGT	9080 ACTCGAGGTT TGAGCTCCAC	9090 CACGGCCGCG GTGCGCGGCG	9100 TTCGCGCGCA AAGAGCGGCT
9110 TCAAGAGGCT AGTTTCCGAG	9120 CGCGGCTTTC AGGTGGAAGG	9130 TCCATCAACT AGGTAGTTGA	9140 CCACCAAGCA GGGTGGTGGC	9150 CCACCAAGCA GGGTGGTGGC	9160 CGGTGCTTCT GCTACATAAC	9170 TCATGTATTC AGTACATAAC	9180 GCTGCGAGCG CCAGCGTCTG	9190 TTCGACCTAA AACGTGGATT	9200 GCACTATATC CGTTGATATC
9210 GGGGTTCGCG CCCCAGGCGC	9220 AGTTCCGCGA TCAAGGCGCT	9230 GGTACCGGAG CCATGGCGTC	9240 CATCTTCAGG GTAGAAGTCC	9250 TGCCGCTTCA ACGGCGAAGT	9260 ACTTTTGTAC TGAAAAATG	9270 CCTCAACGCG GGAGTTGGCG	9280 CGGCTGTGCC GCCGACACGG	9290 AATTGAGGAG TAACTCTCTC	9300 GAGGCTTTCT CTCAGAGAAG
9310 GCTACTCTGA CGGATGAGCT	9320 GCGCGTGTGA CGGGGACAGT	9330 CAGCGCGTGG GTCGCGACCC	9340 AGCGCGAGTT TCGCGCTCAA	9350 TCCGATGTCC AGGCTACAGG	9360 CCGAGAGAGA GGCTCTTCTT	9370 AGAAAGAGTT TCTTCTTCAA	9380 AGAGAGAGAG TCTCTCTTTC	9390 GTATTCGCGG CATAAGGCGC	9400 AGGGGAGAGA TCCGCTTCTT
9410 GAAAGAGAGC CTTCTCTCTG	9420 GCGCGGACCC CGGGGCTGGG	9430 CCTCCGCGCT GGAGGGGCGA	9440 GTGCGCGCGC CACGCGGCGC	9450 TGTGCGCGCG ACGAGCGGCG	9460 TGCGCGTCCG ACCGGAGGCG	9470 CCAGCTGTTT GGTGACAAAA	9480 CGCGAGCTAG GGCTGCTGTC	9490 TAGAGGGGCG ATCTCCGCGC	9500 CCGCTGCGCG GGCGAGGCGG
9510 GTACCAAGCG CATGGTCTCG	9520 CACTGCGCGC GTGAGGCGCG	9530 CCGGCAAGAG GGCGGCTTCT	9540 CGCGCGCGCG CGCGGGGCGC	9550 TCAACTTCTT AGTTGGAAGA	9560 GCGCGCGGCA CGCGCGCGCT	9570 GTACAGGGCG CATGTGCGCG	9580 AATACCAAC TATGCGGTTG	9590 CGCCCCCGCA GCGGGGCGCT	9600 CGGTACGCGC GCCATGCGCG
9610 TCCCTATGCC AGGGATACGG	9620 GCGATGCTGA CGCTAAGCAT	9630 GCTAGAGTTG GCATCTCAAC	9640 TTAAACAACAC AATTGTTGTG	9650 ATCATAGAGG TAGGTACTCC	9660 CGCGCGCTCC GCGCGCGAGG	9670 CTGGACTCGC GACTTGAGCG	9680 TCAGGCTTAG AGTCCGCTAT	9690 CTGGCTAGC GACCGGATCG	9700 CTTTTGGAGA GAAAACCTCT
9710 GCTCTTTTCC CGAGAAAGCG	9720 CAGATTGGTC GTCTAAGCAG	9730 AGTGTACGCG TCACAGTCCG	9740 TTCATCTCGA AAGGTAGGCT	9750 CTCGTGGCAC GAGCACCGTG	9760 CGCGCGCGCG GCGGGCGGCA	9770 CGCGCGCGCG GCGGGCGGCG	9780 CAGCCCCAAC GTGCGGTTTG	9790 AAAGACGCGC TTTCTGGCGG	9800 TCACAGACGA AGGTGCTGCT
9810 CTACTACATT GATGATGTAA	9820 AATTTCATCC TAAAGTAGGG	9830 GCCAGAGTCT CGGTCTTGAG	9840 TGCGGCTTAC ACGGCGGATG	9850 CAGCTGCTTT GTGACAGAAA	9860 CGTGCTACAG GCACCATGTC	9870 GAACCCAGGC CTTGGGTCGG	9880 CGGACGACTT GCTTCTGAAA	9890 ACGCTCTCCG TGCGCGAGCG	9900 CAGCCGATAC GTGCGCATCG
9910 GGGGTCCGAA CCCCAGGCTT	9920 GCAAAACGTG CGTTTTCACA	9930 AGCCGCGTCC TGCGCGCAGG	9940 AGAAACATCA TCTTTGTAGT	9950 TCAGAACGTA AGTCTTGTGAT	9960 CTCGGAAAGA GAGGCTTTCT	9970 TGCGCGTGA ACCGGCACTT	9980 GAAGAGAGGG CTTCTTCTCC	9990 AAGGAGACAA TTCTCTTGT	10000 GGACGTAGAG CTGCTATCTC

FIGURE 23.1 (Continued)

10010 AACGTAGATA TTGATCTAT	10020 GGGAGCGCCG CGCTGCGGCG	10030 CGCCGCTCA GGGCGGAGT	10040 AACCGGGTAC TTGGCCCATG	10050 CACCGCGGGA GTGGCGCCCT	10060 GAAGGAGGGT CTTCTCCCA	10070 ACGCACACTG TGGTGATGAC	10080 GGGCTTCGGG CCCGAAGGCC	10090 GAGTAGCCGA CTCATCGGCT	10100 CTTGGTCCCG GAAGCAGGCG
10110 ATCCAGCCCG TAGGTGCGCG	10120 TGTTGCCGGA ACAACCGGCT	10130 GCCGATTATA CGGCTAATAI	10140 CCGGACGACG GGCCTGCTGC	10150 TGGACGCACT ACCTGCGTGA	10160 CCCATCTGAC GGGTAGACTG	10170 CTTCAGTAGG GAAGTCATCC	10180 TACAGGTGTT ATGTCCACAA	10190 TCGCCACCAT AGCGGTGGTA	10200 ACGCGGGCAC TGCGCCCGTG
10210 AACTACACAA TTGATGCTGT	10220 TTCACGTCGA AAGTGCAGTT	10230 CGGGATTATG GGCCATAACG	10240 CTGGTCAATT GACCACTTAA	10250 GCCAGACCAC TGGTCTGGTG	10260 TGGGCGGACG ACCCGGCTGC	10270 CTCTCGAGCC GAGAGCTCGG	10280 ACATGGACTC TGACTCTGAG	10290 TGGCTCATI AGCGAGTAA	10300 CGGAGCTCA GGCTCTGAGT
10310 GTTTATGCTAT CAATACGTA	10320 CAGCAACGTT GTGCTTGCAA	10330 CAGGCGTGTT GTCCGACCAA	10340 CCATGACCAT GGTACTGGTA	10350 AGGGTGGTTT TCCCAACAAA	10360 TTCAGCGCCG AAGTGGCGCG	10370 CGCGGACCGG GGCGCTGGCG	10380 CATCTCCCGG GTAGAGGGCG	10390 GTCCGATCCC CAGCTAGAGG	10400 ACCGCCCGCG TGGCGGGGCG
10410 AGGCCCCCGG TCCGGGGGCG	10420 TCTAGAAGGT AGATCTTCCA	10430 TGTATTCGGC ACATAAGGCG	10440 TACTATAGGC ATGATATCCG	10450 ATCTACATGG TAGATGTACC	10460 ACCTGTAGGT TGGACATCCA	10470 CCACTACGGC GGTGAATGCG	10480 CGCCGCCACC GGCGCGTGCG	10490 ACCTCCGGCG TGGAGGCGCG	10500 GGCTTTCAGC CGGAAGTCTG
10510 GCCCTGCCCA CGGACGCGGT	10520 AGGTCTACAA TCCAGATGTT	10530 CGCGTCCGCG GGCGAGCGCG	10540 TTTTTCAAGA AAAAAGTGGT	10550 GTATCAAGCC CCATGGTCTG	10560 CTGGGAGACC GACGCTCTCG	10570 GGCCAGTCCG CCGCTCAGCG	10580 CGCGCGTTAG GGCGGCAATC	10590 CAACTCCGAG GTTGACGCTC	10600 ATCTGGCACG TAGACGCTGC
10610 TTTTCTCTCT AAAAGGAGAG	10620 GGACATTGCG CCTGTAAGCG	10630 CCGTAGAAGG GGCACTCTTC	10640 GCACCAAGCC CGTGGTCTCG	10650 ACCTATTATA TGGATAAATT	10660 GCGTTCCTAT CGCAAGGGTA	10670 AGTACCGGCT TCATGGCGGA	10680 GCTGGGCCCA CGACGCGGCT	10690 ACCTCCGGCG TGGAGGCGCG	10700 ATAGCCGGCG TATCCGCGCG
10710 AGGCGGCACT TCCGCGTGA	10720 AGGTACGCGA TCCATGCGGT	10730 ATGGCGGGCG TACCGGCGCG	10740 CACAGCTTGG GTGTGGAACC	10750 GTCCACAGCG CAGGTGTGCG	10760 TGCAGTCTGT ACGTGAGACA	10770 TGCCCCCTCA ACGGGGGAGT	10780 CGAGGAAAC GCTCCCTTTG	10790 CGAAGGAAGG GCTTCCCTTC	10800 TCCGCCCGCG AGCGCGGGCG
10810 CGAGGACGCG GCTGCTGGCG	10820 ATCGAAAAAA TAGCTTTTTT	10830 CCGGTGACCG GGCCACTGGC	10840 GGCGCGCTCG CGCGCGGACG	10850 CATTCGCCAA GTAAAGCGGT	10860 TCCGACCTTT AGGCTGGA	10870 CGCTTTGGTA GCGAAGCAT	10880 ATTACCGAG TAAGTGCTC	10890 CGAGGACAT GCTCCCTGTA	10900 CGCCCTCCCA GCCGAGGGGT
10910 ATAAAGGTT K TATTTTCCAA	10920 CCCAACTCAG GGGTGAGTGC	10930 CGCCCTGGGG GGCGGACCCC	10940 GCCAAGCTCA CGGTTCCAGT	10950 GAGCCTGGGC CTCGGACCGG	10960 GGCCTGACGC CGGACCTGCG	10970 CGCTTGGCCC GCGAACGGGG	10980 CAAAAGGAGG GTTTGCTTCC	10990 GGCAGTACGT CCGTCATGCA	11000 TCTGGGCGCA AGACCCCGCT
11010 ACGTTTAAAG TGCMAATTC	11020 AGGCTTTTGT TCCGAAACAA	11030 CCCTGCTCGG GGGAGAGGCC	11040 GGAAAAAAGC CCTTTTITGG	11050 AAATGGGTCT TTTCCGAGCA	11060 ACGTAGGCCA TGCATCCGCT	11070 CGACGCCGTC GCTCGGCGAG	11080 TACCGGGGGG ATGCGGCCCC	11090 GAGGAGTCTG CTCCTCAGCA	11100 CGCGTTCTCT GCGCAAGAG
11110 GTTCTGCTCG CAAGACGACG	11120 CGCTCTGTAC GGCAGACATG	11130 GTCCGCTGGG CAGGGCACCC	11140 AGGGGAGAGG TCCCTCTCTC	11150 GATGGCGCAG CTACCGGCTC	11160 TCCCTCCCGG AGGAGGGGCG	11170 TGTAGGCGCC ACATCCGCGG	11180 AACTGCGCGC TTCAGCGCGC	11190 TGGCTTACCA AGCATAGGCT	11200 CTAATGCTTG GATTACGAAC
11210 GGGCGCGCGC CCCCGCGCGC	11220 GGCCCGGGCC CCGGCGCGCG	11230 GTGATGGACC CACTACTGCG	11240 TGAACTCTCT ACTTGGAGGA	11250 CCCGCTCCCG GGCGGAGGCG	11260 GACCGCGCGC CTGGCGCGCG	11270 ATCCTCGCGG TAGGAGCGCC	11280 GAGAGGACTC CTCTCTGAG	11290 GCCATGGGTT CGGTACCCAA	11300 CCCCGTCGGA GGGTGACGCT
11310 CTTCCGACTA GAAAGCTGAT	11320 TGGCGACTCC ACGCGTGAGG	11330 GCATGACGCG CGTACGTGCC	11340 CGCCGTCTTG GGCGGAGAAC	11350 GACAAAGCGC CTGTTTTCGG	11360 TGGCGCTCCC ACCGCGAGGG	11370 TCTCTCTGGG AGAGGAGGCC	11380 CTCCTCTACG GAGGAGATCC	11390 CCCTAGCTTT GGGATCGAAA	11400 CAAGGTGCGT GTTCCACGCA
11410 CCCCTGCTCG GGGCGCGAGC	11420 ACCGCGTACC TGGCGCATGG	11430 GGACTTAGCG CCTGAATCGC	11440 CTGCCCAACG GAGCGGTTGC	11450 ACCGGCTCCT TGGCGGAGGA	11460 CTGAAACTTC GGACTTTGAG	11470 GGGCTGCGCG CCCAGCGCGC	11480 CTTGGCCCTA GAACCGGGAT	11490 ATCAGGGGCG TAGTCCGCGC	11500 CGCGGTGTGC CGCGACACGC
11510 ACCAGCGCGC L TGGCGCGCGC	11520 GCTGGACCAT CGACTGDTA	11530 TGGCGTATGC ACCGCATACG	11540 TCGCTGCGCA AGCAGACGGT	11550 CTTGGTCTCT GAACGAGGAG	11560 TAATTGAAGG ATTACTTTTC	11570 TTTTTTCGAA AAAAAGCTT			

FIGURE 23.1 (Continued)

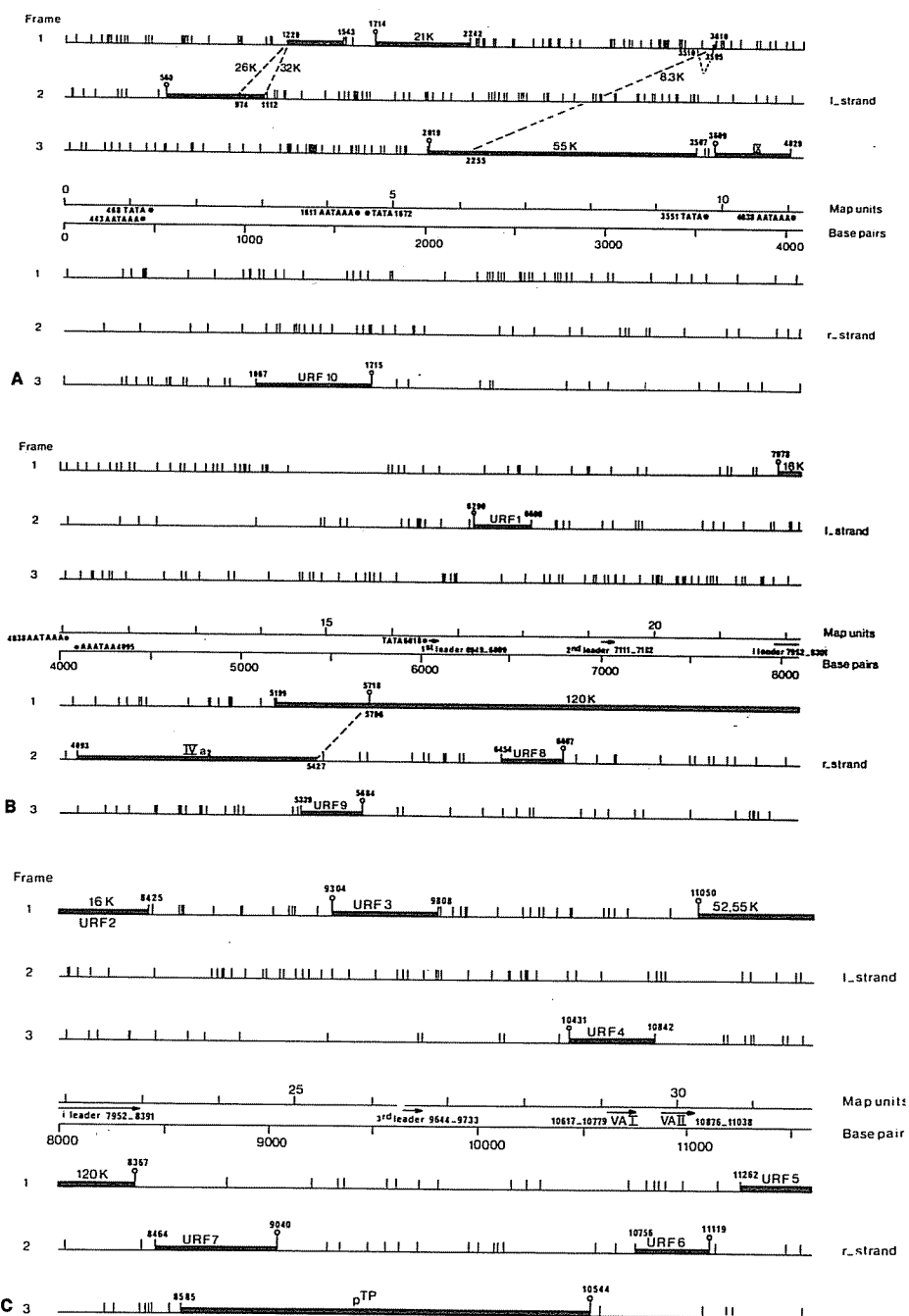


FIGURE 23.2A–C. Structural organization of a region between coordinates 0.0 and 31.7 on the Ad5 genome. This map is derived from the nucleotide sequence in Fig. 23.1. For the positioning of strategic signals, see Fig. 3 (Section VII).

10 20 30 40 50 60 70 80 90 100
 110 120 130 140 150 160 170 180 190 200
 210 220 230 240 250 260 270 280 290 300
 310 320 330 340 350 360 370 380 390 400
 410 420 430 440 450 460 470 480 490 500
 510 520 530 540 550 560 570 580 590 600
 610 620 630 640 650 660 670 680 690 700
 710 720 730 740 750 760 770 780 790 800
 810 820 830 840 850 860 870 880 890 900
 910 920 930 940 950 960 970 980 990 1000
 1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
 1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
 1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
 1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
 1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000

FIGURE 24A-D. Nucleotide sequence of a region between coordinates 59.9 and 71.4 on the Ad5 genome. This sequence and the positions of splice points and leaders were determined by Kruijer *et al.* (1980, 1981, 1983). A schematic presentation of this sequence is shown in Fig. 5 (Section VII).

2010 ACTACGCGCC TGATGGGGG	2020 GCGAGCGCGA GCGTCGGGCT	2030 ACCTCTCTCC TGGGAGAGG	2040 GCGGAGAA GCGCTCTTT	2050 AAGAGAACCC TTCTTCTGG	2060 GCGGTTACCG GCGCAATGG	2070 GTTTAGGCGG CAATCCGCG	2080 GCGCTCCAGC GCGGAGTGG	2090 TACCGCGGCC ATGCGCGGG	2100 CGACCCACAC GCTGGGTGTG
2110 GCGCGGTGGT GCGGCAACCA	2120 GCGGCGAGAC GCGGCTCTTG	2130 ACTACTCAGA TGATGAGTCT	2140 AGGAGCAGGA TCCTGCTCCT	2150 GCGTGAGCTA CGGACTCGAT	2160 TGCGGCGGAG ACGCGCGCTC	2170 TAGGCGAAAA ATCGGCTTTT	2180 AACCCCGCGG TTGGGGGCGC	2190 GGCCCCCTCG CCGGGGAGGC	2200 CGCGCGCTGG GCGCGGACG
2210 CCCTGCCCTT GGGACGGGGA	2220 GCGTGCCAGG CGACACCTCC	2230 AGGTACCAAC TCGATGTTGG	2240 CCCTCTGACG GGGACGCTCG	2250 GCGCGCTGGC GCGCGACGCG	2260 GCGGCGCGGA GCTCCGCGCT	2270 GCCCCACCCA CGGGGCTGGT	2280 AAGCGCGACG TTGCGCTGCG	2290 AGGAGAGGGG TCCTCTTCCC	2300 CTGACCGGTA GACTGGCAAT
2310 AAGGAGAGGG TTCTCTCTCC	2320 ATATCCGTCT TATAGGAGGA	2330 TTTTCTAGTA AAAGATCAT	2340 CCTCAGTCAG GGAGTCAGTC	2350 CTCTTCTTCC GAGAGAAGGG	2360 TGTCGGATTG ACAGCCTAAC	2370 GCGGGGGAGA CGCCCCCTCT	2380 CTCAAGCGGT GAGTTCGCCA	2390 GGTGGCGGAG CCACCGCGTC	2400 GTGGCTACGG CACCGATGCC
2410 CGGTTGGCGG GCCACCGCGC	2420 GATGTTGGAA CTACACCTTT	2430 GGGGCAGCTC CCCGCTGCGAG	2440 CGTGGGGGCG GCACCCCGCG	2450 AACTCTCTCT TTGAGGAGGA	2460 CTTCACTAA GGAAGTGATT	2470 TAGCTCTTCC ATCGAGCGAG	2480 TGGGTCCAAA ACCCAGGTTT	2490 ACATTGCGCT TGTAAAGCAA	2500 CTGCTGCTCC GACGACGAGG
2510 TGGCGAGTCA ACCGCTCAGT	2520 TGGTTGTCTC ACCAACAGAG	2530 CTATTTTTCG GATAAAAAAG	2540 TCTGCTCTCT AAGACAGGGA	2550 GTTGCGCTCT CAACGCGAGG	2560 CGTTTGTCTC GCAACAGGAG	2570 TTGTTGACGC AACAAGTCGG	2580 GCGCCCCCTG CGCGGGGAGC	2590 CTTTCGCTAC GAAAGGCAATG	2600 CGGTGATGGA GCGCTACACT
2610 TCTACACCTT AGATGTGGGA	2620 CTGCTCAGCG GACGACGTGG	2630 ACAACTTCTG TGTGAAGCAA	2640 AGACGCTGCG TCTCAGCGCG	2650 GTACGCTGCT CAGTGCGCCA	2660 AATAGACGCT TTATCTGCGA	2670 GCGCAACGTT CGGTTGCGAA	2680 CTCGCTCTCC GAGCGACGCG	2690 TACAGCGGGA ATGTGCCCCCT	2700 CGCGGTATCG CGCATAGGCG
2710 GTAGAGTCGG GATGTGACGC	2720 AACCGATGCT TTGCTCTAGA	2730 TGGGTTGGAT ACGCGACCTA	2740 AAGAGTGGCG TCTCAGCGCG	2750 CGCATGGGGG GGGTACCGCC	2760 GTTTGGCGTT CAACGCGCAA	2770 CTTTTGGCGT GAAACGCGCA	2780 GTACGCTGCG CATGCGAGCC	2790 GTTGGGCGCG CAACCGCGCG	2800 GAGTTGAAGA CTCACTTCTT
2810 TGGGCGATAA ACCCGCTATT	2820 ACCGGACGGT TGCGCTGCCA	2830 CTCCACGAAAC GAGGTGCTTG	2840 GGTGGATAGT CCACCTATCA	2850 GTAGAAAAAG CATCTTTTTC	2860 GTTTGGAGCT CAAACTGCGA	2870 TCTATGGGGA AGATACCCCT	2880 TAGGACGCGA ATCCGCTCGT	2890 CGGTTGGCGT GCCAACCGCA	2900 CGGCTGCGCT GCCGACGCGA
2910 CTTGCTGCGC CAAGGACGCT	2920 GCGAGCGCGG GCTTGGCGCG	2930 TCCCGCGGCA AGGCGCTGCT	2940 GTATGCGACA CATACTGAT	2950 TACCGGAGCG ATCGCTCGCG	2960 AGTTGCTTCA TCACCAAGAT	2970 CGGTTTTTGA GCCAAAATC	2980 AAACTCCCAAG TTTGAAGGTC	2990 AACCTGCGCA TTGGACGCGA	3000 CGCTCTGCGG CGAGAGGCGC
3010 GCGCGTTTGG GCGGAAAGC	3020 GAGACGTTGT CTCTGACACA	3030 CCTTTTGTGG GGAAGACAGC	3040 CTTTTACTTT GAAATGAAAG	3050 CAGTGAGACC GTCACTCTGG	3060 TCACAAACAC AGTGTGGTG	3070 CTTGAGCTCC GAACCTCGAG	3080 CACTGTGTGG GTGACACGCG	3090 GCGGATCTGG GCGGCTAGCC	3100 CATGATTTTG GTACTAAAGC
3110 CGTGAGTCTG GCGCATCGA	3120 CCAGTGGGTG GGTCAACCCAC	3130 AAACGGATGG TTTGCTTACC	3140 GCGGCTGATT GCGCACTTAA	3150 GGATGGGGGG CTCAACCCCGC	3160 TTCCAGTACT AAGGTCAATG	3170 CGTGTGAGTA GCAAGTCAAT	3180 CTCACTGACG GAGTGAAGTG	3190 TAGCAGCGGG ATCGTGGCGC	3200 CAGCGCTGCG GTGCGGAGCC
3210 GGACCTCTCC CCTGGAGAGG	3220 CTACGTTTAA GATGCAAAAT	3230 ACGTTCTTGT TGCAAGAACCA	3240 TTGTCTCTCT AACAGAGGAG	3250 CCGGAATGGC GGCTTACCGC	3260 GTCAACGCGT CAGTTGGCGA	3270 GCTGCTGCGT CGAGCAGCTA	3280 CGCGGACCTG GCGGCTGGCG	3290 AAGTTTGGCG TTCAACGCGC	3300 GCTCGGACGG CGAGCTGGCC
3310 CTGAACCTCC GACTTGGAGG	3320 TCGCTGCGTT AGCGACGCAA	3330 TGATTACTAC ACTAATGATG	3340 CGGCGTACGG GCGCGAGTGC	3350 AGCAATGGCA TCGTTACCGT	3360 CCTCGAAGTC GAGCTTGGAG	3370 ACGTACGCTG TGATGCGAGC	3380 CCAAGAAACG GGTCTTTTGG	3390 ACTGGGCGCT TGACCGGAGG	3400 TAGCTGCGGT ATGCAAGCGCA
3410 TGGATCTCTT AGCTAGAGGA	3420 TTGTAGCTGG AACATTGGAC	3430 ATGTGGAAAG TACACCTTTC	3440 CTGTCCCGAT GACAGGGCTA	3450 GCATGCGGTC GCTACGCGAG	3460 GCGACGTTCT GCTTCAACGT	3470 AGAGGTTGCA TCTCAACGCT	3480 CCTCGAGACG GGAGCTCTGC	3490 TTGACGCGGA AACCTGGTCT	3500 GGATGGAAAC CCTACTTGGG
3510 TTAAACGCTG AATTTTGGAC	3520 CTTTTGGCGG GAAAGCGCGC	3530 AACCGGTTTT TTGGGCAAAA	3540 GCACGAAGTA CGTGTCTTAT	3550 AGGTGCGAGT TCCAGCTCTA	3560 TCCCGCTCCG AGGCGGAGGC	3570 GCGGCGCGCG GCGGCGCGCG	3580 ATGCGAGCGG TAGCTGCGCG	3590 TGACGCAAAI ACTGCGTTTA	3600 GAATAAAGAT CTTATTTCTA
3610 ACGATGTGGA TGGTAGACCT	3620 CGGCTTGGCG GGCAGACGCG	3630 GTACCGGCAA CATGGGCTTT	3640 ACCGTGTCTA TGCGACAGCT	3650 CGAACCTCTT GCTTGAGGGA	3660 CAGCTTGGAG GTGCAAGCTC	3670 TTCTCTGAGC AAGGACGCTG	3680 TCTTTGACGA AGAACTGCTT	3690 TTTCTGTTTT AAGCAAAAC	3700 AACTTCTGCG TTGAAGGACC
3710 ATACCTGCGG TATGGAGCGG	3720 GAAGTTGCTC CTTCAAGGAG	3730 GCGAGGCAAC CGCTCGGTGG	3740 GGCGCGTGGG CGCGGACACT	3750 CGCGCTGTAG GGCGGACATC	3760 TAAAGGGGCG ATTTTCCCGG	3770 TTGCGGACGA AACCGCTGCT	3780 ATTTTGGGAC TAAACCCCTG	3790 GTTGTCCGAC CAACAGGCTG	3800 ACCGTCTGAA TGCCAGACTT
3810 GTGGTCAGTT CACCAGTCAA	3820 TGCTACAACG AGCATTTTGG	3830 TCTTGAATAG AGAATTATAG	3840 CTTGAATATG GAATTTTATC	3850 GATCTGCGGA CTAGAGCGCT	3860 GTCTTGAAGA CAGGAATCTT	3870 GCGGCGGTGG GCGCGGCAAC	3880 ACGACAGCTG TGCTGTGCAAC	3890 AAGGATCGCT TTCTTAGCGA	3900 GAAACAGGGA CTTTGTGCCC
3910 TAATTATGAC ATTAAGTACG	3920 CGCTTACGCG GCGAATGCGC	3930 AGGCGGCGAA TCCGCGCTTT	3940 ACCGCGGTGA TGGGCGCACT	3950 CGATGGAAGA GCTACTTCTT	3960 CGTGGATGCG GAGCTAGGCG	3970 TTGATGGAAC AACTACCTTG	3980 GGATGCTGAG CCTACACTTC	3990 ACTGTATATC TGACATATG	4000 CTTCTGCACT GAAGACGCTA
4010 CGCACTGCGC GCGTGACGCG	4020 AGATGACATC TCTACTGAGG	4030 ACAGTGAACG TGCTACTGTC	4040 CGACGTTGGA GCTCGAAGCT	4050 TACGTGGGCG ATCGACCGCG	4060 GTGCGGAGGG CAGCCGCTCC	4070 ACCAACGCTT TGTTTGCGAA	4080 AAGCGTCGAC TTGCGAGCTG	4090 GAATTCGTTT CTTAACGAAA	4100 CAGTTTAATA GCTCAATATAT
4110 GCCATGGG D CGGTACC									

FIGURE 24 (Continued)

10 20 30 40 50 60 70 80 90 100
 15 3' GGAACTGTTT TCTTGGGTGT GACTAATACT GTGCGTATGA GCCTCGATAC GATTGGTCCG ATCGGGGCTA CATTCGAAC AACGTACCGG CCGCTAATAT
 15 CTTTGACAAA AGAACCCACA CTAATTATGA CACGCTACT CCGAGCTATG TAGCCCGCAT GTAAGCTTTG TTGCTAGGCG GGTGA*AA*
 110 120 130 140 150 160 170 180 190 200
 TTACGTTCCA CGACGAGTTT TTTAGTCCGT TCGGAGGCG GTTTTTCCTT TCGGTGAGCA TCAGTACGAG YACGTCTATT TCCGTCCATT CGAGGCTTTG
 AATGCAAGGT GGTGCTCAA AAATCAGGCA AAGCCTCGCG CAATAAAGAA AGCACATCGT AGTCATGCTC ATGCAGATAA AGGAGGTTAA GGTCTGGAA*
 210 220 230 240 250 260 270 280 290 300
 GTGGTGTCTT TTTCTGTGGT AAAAGAGAG TTTGTACAGA CCGCCCAAGA CGTATTTCGT GCATAAACAC AAAATAAAT TTGTTTTTTT GTAAATTTGT AATCTTCGGA
 CACCACAGAA AAAGACACCA TTTTCTCTC AAACATGTCT GCGGGTTTCT GCATAAACAC AAAATAAAT AACAAAAAAA CATTTTAAACA TTAGAAGCTT
 310 320 330 340 350 360 370 380 390 400
 CAGAAATGTTG TCCTTTTGTG TGGGAATATT CGTATTCTGC CTGATGCCGG TACGGCCGCA CTGGCATTTT TTGGACCAAT GGCCTAATTT TTAGGAGCTT
 GCTCTCAAC AGGAAAAACA ACCCTTATAA GCATAAGAGC GATAAGAGC GACTACGGCC TACGGCCGCA CTGGCATTTT AAACCTGGCA CCGTATTAA TTAGGAGCTT
 410 420 430 440 450 460 470 480 490 500
 GCTCTCAAGG AGCCAGTACA GGCCTCAGTA TTAGCTCTGC AGCCATTGCT GTAGTCCAAC TAAGTAGCCA GTCACGATTT TTCGCTGGCT TTAGGAGCTT
 CCAAGCTCC CCGGTCTATG TCGGTATGAT AATGTAAAGC TCGGTAAACA CATCAGGTTG ATTCATGCTG CAGTGTAAA AAGCAGCCA AATAGCTCCG
 510 520 530 540 550 560 570 580 590 600
 CCGTATTGTA TGGCGTCCG CATCTCTGTT GTAATGTCGG GGTATCTCTC CATATTGTTT TAATTATCTT CTCCTTTTGT GTATTGTGG ACTTTTTGGG
 GGAATAACAT ACCCGCAGGC GTAGACACAA CATTACAGCC CCGTAGAGAG GTATAACAAA ATTAATAGGA GAGAAAAACA CATATAACCC TGAATAACCT
 610 620 630 640 650 660 670 680 690 700
 AGGACGAGTC CGTTTATGCG TGGGAGGCG AGGCTCTGTT GTATGTCGCG AAGTGTCCGC GTCCGATTGT CAGTCGGAAT GGTCTATTIT TCTTTTGGAT
 TCTTGGCTAG GCAAAATAGC ACCCTCCCGC TCCAGAACAA CACACAGCGC TACACAGCGG TACACAGCGG CAGCTTAACA CAGCTCCTTA CCAGTAAAAA AGAAAACTTA
 710 720 730 740 750 760 770 780 790 800
 AATTTTTTIG TGGTACGATG TCCGCTGGTC GAGTTAGTCA GTGTACGATT TTTTCCCGGT TCACGCTCTC CTCATATATA TCTTGATTTT TTACTGCTAT
 TTAATAAAC ACCACTCGAC AGGCACACAG CTEAATCAGT CACAGTGTAA AAAAGGCGCA AGTGTCTGAG AGTCAGAGC GAGTATATAT AGGCTAAAAA AATGAGCTAA
 810 820 830 840 850 860 870 880 890 900
 GCCAATTTCG GGTGTTTTTT GTGGGTCTTT TGGGCTGCGG TTGGATGCGG GTCTTTGCTT TCGGTTTTTT GGTGTGTGAA GGAGTTTACG AGTGAAGGCA
 CGGTAAAGT CCACAAAAAA CACCCAGAAA ACCCGACCGG AACCTAGCGC CAGAACGAAA AGCCAAAAAA CCCACAACCT CECTCAATCG TCACTTCCGT
 910 920 930 940 950 960 970 980 990 1000
 AAAGGGTGCA ATGCATTGAA GGGTAAATT CTTTGTATGT TAAGGTTTGT GTATGTTCAA TGAGGCGGGA TTTTGGATGC AGTGGGCGGG GCAAGGGTGT
 TTTCCACGCT TACGTAACCT CCAATTITTA GAAACTTACA ATTCCCAACA CATACAAGTT ACTCCGCCCT AAAGCTTACG TCACCCGCGC CGTCCCAAG
 1010 1020 1030 1040 1050 1060 1070 1080
 GGGGCGCGGT GAGTGTGTTG AGGTGGGGGA GTAAATATAT AACCAAGTT AGGTTTTTAT CCAATATAATA ACTACTAC
 CCGCGCGCCA CGTCACAAAC TCAACCCCTT CATTATCATA TTGGCTTCAA TCCAAAATAA GGTATATTAT TGATGATG

FIGURE 25.1. Nucleotide sequence of a region between coordinates 97.0 and 100.0 on the Ad5 genome. This sequence was determined by Steenbergh *et al.* (1977) and Steenbergh and Sussenbach (1979). The strategic sequences were determined by Baker and Ziff (1980, 1981) and further derived from the Ad2 sequence of this region (Fig. 21).

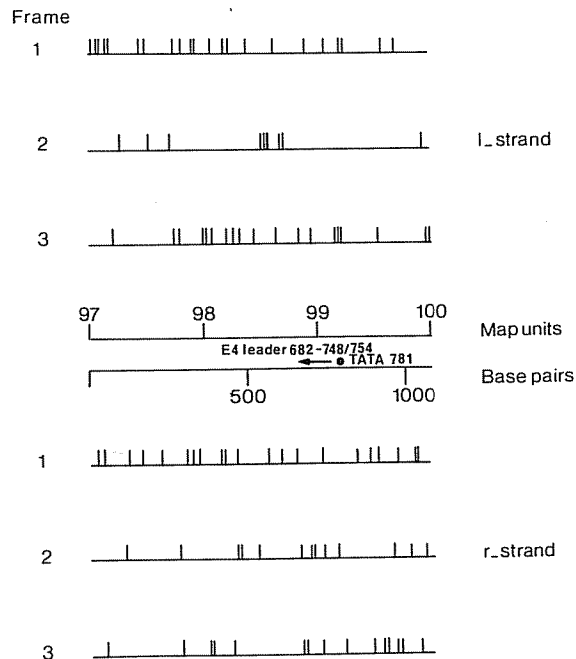


FIGURE 25.2. Structural organization of a region between coordinates 97.0 and 100.0 on the Ad5 genome. This map is derived from the nucleotide sequence in Fig. 25.1.

10	20	30	40	50	60	70	80	90	100
15	25	35	45	55	65	75	85	95	105
110	120	130	140	150	160	170	180	190	200
210	220	230	240	250	260	270	280	290	300
310	320	330	340	350	360	370	380	390	400
410	420	430	440	450	460	470	480	490	500
510	520	530	540	550	560	570	580	590	600
610	620	630	640	650	660	670	680	690	700
710	720	730	740	750	760	770	780	790	800
810	820	830	840	850	860	870	880	890	900
910	920	930	940	950	960	970	980	990	1000
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000

FIGURE 26.1A–K. Nucleotide sequence of a region between coordinates 0 and 31.7 on the Ad7 genome. This sequence and the positions of strategic sequences were established by Dijkema and Dekker (1979), Dijkema *et al.* (1980a,b, 1981, 1982), van Beveren *et al.* (1981), Engler (1981), Engler *et al.* (1981, 1983), and Engler and van Bree (1982).

2010 TAGAATCTAA ATCTTAGATT	2020 TGACCGGTCA ACTGCGCAGT	2030 CGTCGGAGAC GCAGCTCTCT	2040 CCTCATCGTC GGAGTAGCAG	2050 CCTATGACTC GGATACTGAG	2060 TGTGGGTGGC ACACCCACCG	2070 TGGTACGGTC ACCATGCCAG	2080 GCCAAGACGT CGGTCTGCA	2090 CCTCCTCGTC GGAGGAGCAG	2100 GTCCCTCCTGT CAGGAGGACA
2110 TAGGCTCTCG ATCCGAGAGC	2120 GCCGGACCTG CGGCTCGGAC	2130 GGAGGCCACC CCTCCGCTGG	2140 TCCTCATCGA AGGAGTAGCT	2150 CTGGACAAG GACCTGTTC	2160 GACTTGACGC CTGAATCGCG	2170 TGCCCAAGAA ACGGGTGCTT	2180 TGATCCAGAT ACTAGGTCTA	2190 GCTGGTCACC CGACCACTGG	2200 TGCTCTGTCC ACGAAACAGG
2210 CCTTAATTCT GGAAATTAAGA	2220 CCCTCTCCTT GGGAGAGGAA	2230 AGGATCACCC TCCTAGTGGG	2240 TTATTAAGTT ATAATTCAA	2250 CTTGGCTCAA GAACCGAGTT	2260 CGGAATTCA GGCTTTAAGT	2270 AATTACTCGG TTAATGAGCC	2280 CGTCCGAGGG GCAGGCCCTC	2290 ACTTTGACAA TGAACTGTT	2300 ACCACCGTAC TGGTGGCATG
2310 TCCAAGTCTC AGGTTCAAG	2320 GCTTCCGTC CGAGGCGAGG	2330 CTACTACAAA GATGAAGTTT	2340 GTATAACGT CAATATTGCA	2350 CCTCTTTATA GGAGAAATAT	2360 AGTAATCTTG TCACTAGAAC	2370 TTGAATCTGT AACTTAAGAC	2380 GACAACCAAC CTGTTGGTTG	2390 CTTGGACTCC GAACCTGAGG	2400 TACTAACCTT ATGATGGGA
2410 CCACCGGTAA GGTGGCCATT	2420 TCCTTAATAC AGGAATATG	2430 GATTCATAG CTAAGATATC	2440 AGACTCCGGA TCTGAGGCTT	2450 CTATTGTGTA GATAAACAA	2460 TATCTTAATG ATAGAATTAC	2470 ATTCCTCTAA TAAGAAGATT	2480 TTATAATCTT AATATTAGAA	2490 TACGTACGAT ATGATGGGA	2500 GTATAGTCCC CATATCAGGG
2510 TTACCCGCTC AATGGGCGAG	2520 TCCAATATTA AGGTTAATAT	2530 TCTATGTGTT AGATACACAA	2540 CTATTGTGCT GATAAAGCAG	2550 GAAAATCTAC CTTTTAGATG	2560 AACATACACT TTGTATGATG	2570 CCATACACCG GGTATGTGGC	2580 GTCCCAACAA CAGGGGTTGT	2590 GCCGTACTTT CGCGATGGAA	2600 CGTTATGTGG GCAATACAC
2610 AATAGTTATA TTATGAATAT	2620 ATCCAAATCT TAGGTTAAGA	2630 CCCTTACCCA GGGATGGGTT	2640 TATTACCGTA ATAATGGCAT	2650 ACATAAATAC TGTTATTATG	2660 CGATTGTGAT GCTAACACTA	2670 TGCACTAAGA AGCTGATCTT	2680 TGACCAACAA ACATGTTGTT	2690 TCGAAAACAA AGCTTTTGTG	2700 CCAAATTAAT GGTTTATATA
2710 ATGCACACAT TACGTGTGTA	2720 CTTCGAACCC GGAAGTGGGG	2730 CCGTTCATTC GGCAAGTAG	2740 ACACTCCCCA TGTGAGGGTT	2750 ACATCAAAAC TGAGTTTITT	2760 TACGTACGAC ATGCACTGCT	2770 CTAACGTTGT GATTGCAACA	2780 AGTCCATCCC TCAGGTAGGG	2790 ACTTCTCAGT TGAAGAGTCA	2800 CAACAGACAC GTGTCTGTGG
2810 TTCTTTACGT AAGAAATGCA	2820 ACAACTCTCT TGTGTTAGAG	2830 TACATTAGAA ATGTATCTTT	2840 CCGTATGACT GGCATACTGA	2850 TACTTCCACT ATGAAGGTGA	2860 TCGTTCCGAG AGCAAGGGTC	2870 GGGGTACGCG CGCCACTGCG	2880 GTGGATGTCT CAGGTACAGA	2890 TTGACGGACG AACTGCCCTC	2900 AAGTAAGATT TTCACTCTAA
2910 ATTTCCCTTT TAAAGGGA	2920 ACGGTGCACAC TGCAGTGTGG	2930 TTGCTATTAT AAGCATATA	2940 ACTAGACACC TGATCTGTGG	2950 TGTAAGCCTA ACATTGCGAT	2960 CTCTCCGGAA GAGAGGCCCT	2970 TAGTCTACGA ATCAGATGCT	2980 TTGGACGCGA AACCTGCCCT	2990 CCACCTGTAA GGTGACATT	3000 CGTTATAAGA GCAATATCTT
3010 ACGATGGCAG TGCTACCGTG	3020 GTATAGCAAA CATATCGTTT	3030 GTGTACGTGC CACATCGACG	3040 GTTCCTTACC CAGAAATGCG	3050 GGACATAAAC CCTGTATTGT	3060 TGTATTACAA ACATATATGT	3070 CTAATGCTTC GATTACCAAG	3080 ACGTGGTACG TGCAACTGCG	3090 TATATCCACC ATATAGGTGG	3100 AGCCCTCCCT TCCGAGGGGA
3110 TACAATACG ATGTTTATGC	3120 GAATGGTCAC CTTACCACTG	3130 ATTGTACTTA TAACATGAAT	3140 GTCACTTCCC CATGTGAAGG	3150 ATTACAACCT TAATGTTGGA	3160 TGGTCTACGG ACCAGATGCC	3170 AAAAGGTCTC TTTTCCAGAG	3180 ACTCGCATTC TGAGCGTAA	3190 TCCCTAGAAA AGGAATCTTT	3200 CTATACTTAT GATATGAATA
3210 AAGTTGATAC TTCAACTATG	3220 CTTCTAGGAC GAAGATCTCT	3230 TCTATCTAC AGATATGATG	3240 TGTGATTGG ACACTAAACC	3250 TTCCCAACCG AAGGGTCCCG	3260 CGTACGCTTA GCATGCCAAT	3270 CGCTCTCGTT GCGGAGGCAA	3280 CGTACGATCT GCATGCTAGA	3290 AAGGTCCGCC TTCCAGCCGG	3300 ACACGGACCT TGTGCGTGGA
3310 ACACTGACTT TGTACTGAA	3320 CTGGACTCCG GACCTGAGGC	3330 GGCTAGTAAA CGATCATATT	3340 CCACGAACGG GGTCTTGCC	3350 ACGTGACCTC TGCACGTGAG	3360 GCCCAAGGCC CGAGTTCGG	3370 AAGATCACCA TTCTAGTGGT	3380 CTTCTTTGAC GAAGAACTG	3390 TGATTTCATT CTATGGGGG	3400 CATCACCCCC GTATGGGGG
3410 GTTTACACCC CAAAATGTGG	3420 TACCCCTGAA ATGGGGACTT	3430 AGTCCAAACA TCAGTTGGT	3440 TTCACCTGCT AAGGTGGACA	3450 TTAACCAATT AATTGGGTAA	3460 TAAACAATT ATTCTGTCT	3470 AAAGACAGAA TTCTGTCTCT	3480 CGTGCAGGCT GCAGCTGCEA	3490 ACTCACCTTC TGAGTGGAG	3500 GCGAAGAAAA CGCTCTCTTT
3510 CTCCCCCTTC GAGGGGGGAG	3520 ATAAATCGGG TATTTAGCCC	3530 AATAGACTGC TTATCTGAGC	3540 CCGTCCGAGG GGCAGGCTCC	3550 GTGGTACCCG CACCATGGGC	3560 TCCCAAGCAA AGGAGTTCGT	3570 GTCTTACAGT CAGAATGTCA	3580 ACCTTAGGTC TGGATCCAC	3590 ACACCTACCC TGTGATGGG	3600 TGTGGCGAGG AGACCCGTTC
3610 TCGGGCGGTT AGCCGCCCAA	3620 AAGGAGTTGC TTCTCAACG	3630 GACTGGATAC CTGACTATG	3640 GGTGAACATC CCACTTTGAG	3650 AAGCACTGGT TTGCTCACCA	3660 AACCTACGTC TTGGATGCAG	3670 GACGTGGGCG CTGACGCCCG	3680 GCGGCGATGA CGCCGCTACT	3690 CGACGGCGGT GCTGCCGCCA	3700 TGTGGTAGGA ACACCATCTT
3710 ACCTTACCGG TGGATGGGCG	3720 ATAATGCCCT TATTACGGAA	3730 CGTAACAACG GCATTGTGCG	3740 GTTAAGGTCA CAATTCACAT	3750 AGGAGATTAT TCTCTAATA	3760 TAGGAAGTTG ATCCTTCAC	3770 GGACCGACTC CCTGGCTGAG	3780 CTGTTCTGAG GACAACTAC	3790 AACAAGAGAA TTGTTCTCTT	3800 CCGAGTCGAG GGCTCAGCTC
3810 CTCCGGGATT GAGGCTCTAA	3820 GGGTTGGGAA CCCAACGCTT	3830 TCCGCTTGAC AGGGGAACTG	3840 AGATTCCGTC TCTAAGCAGG	3850 ACCGGGTCAA TGCCCGCAGT	3860 CGCACTCGTT GCGTGAGCAA	3870 TGACTCAGAC ACTGAGTCTG	3880 GACAACGGTG CTGTTGCCAC	3890 TCGTTTCAGA AGCAAGTCTT	3900 TTTATTTCTA AAATAAAGAT
3910 GAGTTTAGTT DCTCAATCAA	3920 ATTATTTCTT TAAATAAAGA	3930 TTATGACCAA AATACTGTTT	3940 TATTTTGTGT TACTTACAAA	3950 TAACTAAAA ATGATGTTT	3960 ATTTGATTTT	3970 AGCGCGGCGC TCGCGCGCGG	3980 ATACGGGACG TATGCCCTGG	3990 TGGTAGGCAA ACCATCGGTT	4000 AGCTAGTAGC TCGATCATTG

FIGURE 26.1 (Continued)

4010	4020	4030	4040	4050	4060	4070	4080	4090	4100
TCTTGAAGCA	CCTAGAGGAG	GTGATGGGAC	ATTTCACCCC	TAACTTACAA	ATCTATGTAC	CCGTAATCAG	GCAGAGCCCC	CACCTCTACT	GAGGTAACTT
AGACCTCGGT	GGATCTTTTC	CAGTACCCTG	TAAAGGTGGG	ATTGAAATGT	TAGATACATG	GGCATTAGTC	GGTCTCGGGG	GTGGAGATG*	CCTCATTTGA
4110	4120	4130	4140	4150	4160	4170	4180	4190	4200
CTCGGAGAAC	GAGGCCCCAT	CACAATATTT	AGTGGGTGAG	TATCGTTTCA	GCCTCACGTA	CCACACAGTG	TTATAGAAAA	TCCTCGCTTG	ATTAACGTG
GAGCCTCTTG	CTCCGGGGTA	GTGTTATAAA	TCACCCAGTC	ATAGCAAGGT	CGGAGTGCT	GGTGTGTGAC	AATATCTTTT	AGGAGCAGAC	TAATTCACAC
4210	4220	4230	4240	4250	4260	4270	4280	4290	4300
CCCCCTCGGG	AATCACATCC	ACAAATGTTT	AGACAACTCG	ACCCCTGCCA	CGTAGGCCCC	ACTTTAATAT	ACGTAAACCC	TGACCTAGAA	CTCCAAACCG
GGGAGGCCCC	TGATGTAGG	TGTTTACAAA	TCTGTGAGC	TGGGACGGGT	GCATCCGGGG	TGAATTATTA	TGCAATTTTG	ACTGATCTTT	GAGGTGGGCA
4310	4320	4330	4340	4350	4360	4370	4380	4390	4400
TACCAAGCGG	GATCTAGGGC	AGAGCCCAAG	TATAACACGT	CCTGGTGGTT	CTGTACACATA	GGCCACGTGA	ACCCCTTAGA	TAGTACGTG	AATCTCCCTT
ATGTTGCCCG	CTGATCCCG	TCTCGGGTTC	ATATTGTGCA	GGACCAACAA	GACAGTGAT	CCGGTGCACT	TGGGAATCT	ATCTCCCTAC	TTAGAGGGAA
4410	4420	4430	4440	4450	4460	4470	4480	4490	4500
TCGTAACCTT	TTTAAACCTC	TCCGGAACAA	CTGGGGGTTC	TAGAGGCTAC	GTGAGTAGGT	ATTACTATCG	CTACCCCGCC	ACCCCTCGCC	GTGCCCCGTT
AAGCATGAAA	AAATTTGAG	ACGCCCTTGT	GACCCCGCAG	ATTCTCCATG	CACCTACCCA	TAATGATAGC	GATGGGCGCG	TGGGCGCGCG	CACGGGCGAA
4510	4520	4530	4540	4550	4560	4570	4580	4590	4600
ATGTAAGCGG	CCTAGTATTT	GCAGTATCAA	CACAAGGTCC	TACTCTAGCA	GTATCCGGTA	AAAATGTTTG	AAACCCCGCT	CCACCGGTCT	AACCCCTTAC
CAGCTGCCGG	CGTACATATA	CGTACATATA	GTGTTCCAGG	ATGAGATCGT	CATAGGCCAT	TTTTCACAA	TTTGGCGGGA	GGGTGCGAGA	TTGGGGGATG
4610	4620	4630	4640	4650	4660	4670	4680	4690	4700
TTTCAAGGGA	CACCGGGGCC	TCGTATACAA	GGAGTGTAT	AAAGCTAAG	GGTCCGAAAG	TCAAGCTCTC	CCCCCTAGTA	CAGGTAGCTA	CCCCGATAT
AAAGTTCCTT	GTGGCCCGGG	AGCATAGTAT	CCCTCACATA	TTTGCAATTC	CCAGGCTTTC	AGTTCAAGAG	GGGGGATCAT	GTCCACATGC	GGGGCTATAA
4710	4720	4730	4740	4750	4760	4770	4780	4790	4800
TTTATGGGCA	AAGACCTCGG	CCCCACTAAT	TGACCTTACT	CTCGTTTAAG	GATTCTGCGA	CTCTGAACGG	CGTGGGCGAC	CCTGGGCTAT	ACTGGGGGTA
AAAATACCGT	TTCTGGAGCC	GGGGTATAT	CTGGGATGA	GAGCAATTT*	AGCAGCT	GAGACTTGCC	GCACCCCGTG	GGACCTTAAA	TGACCCCAAT
4810	4820	4830	4840	4850	4860	4870	4880	4890	4900
ATGCCCAACG	TCTACATCA	AATCCCTCGC	TGTCGAGGCG	AGGAGGGGCT	CGTCCCGCCG	GTGAAGCAAG	TAGTAAAGGG	AATGTACTCT	TAAAGGGGCG
TACGGGTGTC	AGATGGTAGT	TTAGGGAGCG	ACAGCTGCGG	TCCTCCCGGA	GCAGGGGGGG	CACCTGCTTC	ATCATTTCCC	TTACATGGAT	ATTTCCTCCG
4910	4920	4930	4940	4950	4960	4970	4980	4990	5000
TGGTTCCGGC	AATCTCCCGC	GAGAGGGGGT	TCACATCTCT	CGAGGACCTC	GCCTCTCTTC	AAAAGTCCGC	CGAAGTCCGG	CAGTCCGGTAC	CCGTAAACCC
E ACCAATCCG	TTAGGAGGGC	CTCTCCCGCA	AGTGATAGAA	GTCTCTGGAG	CGAGGAGAA	TTTTTCAGG	GTCTCAGCCC	GTACAGCCAT	GGCATTTTGG
5010	5020	5030	5040	5050	5060	5070	5080	5090	5100
TTTCTCAGAC	AAGCTTCTCG	AGCTCGGCCA	GGGCTCCGAG	CCACTACACG	AGATACCTGA	GAGCTAGGTC	GTCTGGAGGA	GCAAAAGGCC	CAACCTCGCC
AAAGAGCTCG	TTGCAAGAGC	TCGAGCCGGT	CCGAGAGCTC	GGTGATGTGC	TCTATGGCAT	CTCGATCCAG	CAGACCTCTT	CGTTTCCGCG	GTGGGGACGG
5110	5120	5130	5140	5150	5160	5170	5180	5190	5200
GAGGACCTCA	TCCCTTAGTC	TGCTACCCGC	AGTCTGCCAG	GGTCCCGAGC	TAGGAAGGTA	CCAGCGTCCG	AGGCTCAGTC	CCAACAAGGG	CAGTGGCACT
CTCTCGAGGT	AGGGAATCAG	ACGATGGGCG	TCAGGCGCTG	CCAGGCTCCG	ATCCTTCCAT	GGTCCGACGG	TCCGAGTCAG	GGTGTGTTCC	GTACGGGTGA
5210	5220	5230	5240	5250	5260	5270	5280	5290	5300
TCCCAACCGG	CGGACCAACC	CGCGAACGCT	CCACGCGGAA	GTCTGAGTAG	GACGACCAAC	TCTTGGCGAC	GGCTAGCCCG	GGGAGGTACA	GGCGGTCCAT
AGGGGTGGCG	GGCTGGTTGG	GGCGTTGGCA	GGGTGCGCTT	CAGACTCATC	CTGCTGGTCC	AGAACCCTGG	CCGATCGGCG	CCCTGCATGT	CGGCCAGGTA
5310	5320	5330	5340	5350	5360	5370	5380	5390	5400
CGTCAATAGG	TATTCAGGCA	TCAACTCGCG	GAGCCGGGCG	ACCGGAACCC	GTGCTCTGAA	TGGAAACCTT	CAAAATACCG	TCCGTCCCGT	CATCTATGTA
GCAGTTTACC	ATAGATTCGT	AGTTGAGCGC	CTCGGCGGCG	TGGCCTTTGG	CACGGAGCTT	ACCTTTGGAA	GTTTTATGCG	AGGCAGGGCA	GTAGATACAT
5410	5420	5430	5440	5450	5460	5470	5480	5490	5500
AACCTCCGTA	TGTGGAACCC	GGGCTCCTTT	TACCTAAGCC	CCCTCATACG	TAGGCGTGCC	GTCCCTCTCG	TCTGCGAAG	CGTGAGTGT	TCGGTCCAGT
TTGAGGGCAT	ACAGCTTGGG	CGCGAGGAAA	ATGGATTCCG	GGGAGTAGCC	ATCCGACCGC	CAGGAGACGC	AGACGGTTTC	GCATCCCA	AGCCAGGTCA
5510	5520	5530	5540	5550	5560	5570	5580	5590	5600
CTAGGCGCAG	TAGTCCCACT	TTTTGTTCAA	AAGGCGGTAC	AAAAAAGTAC	GCAAGGAATG	GAAACCAAG	GTACTCAAGC	ACAGGTGCGA	CCCACTGTTT
GATCCGGCTC	ATCAGGGTCA	AAACAAGTT	TTCGCCATG	TTTTTTGATG	CGTTCTTAC	CITTTGTTTC	CATGATTTTC	TGTCCACGCT	GGGTGACAAA
5610	5620	5630	5640	5650	5660	5670	5680	5690	5700
CTCCGACAGG	CACAGGGGCA	TCTGGCTGAA	ATACCCGGAC	AGGAGCTGCG	CTCACCGAGC	CAGGAGAGCG	ATCTCTTAG	GTGCGGTGAG	ACTATGTTT
GAGGTGTGTC	GTGTCCTGCT	AGACGACATT	TATGGGCTCG	TCCTCGAGCG	GAGTGCTCG	GTCTCTTTCG	TAGAGGAATC	CAGCCCACTC	TGATACAAA
5710	5720	5730	5740	5750	5760	5770	5780	5790	5800
CGCGACAGGG	TCCGGTCCGT	TTTCTCCGCG	TGACCTCTCC	CCATCGCCAG	CAACAGTTGG	TCCCTTAGGT	GGAAAGATG	CCATACATT	GTGTACAGGG
CGCGGTGTCC	AGGCGACAGC	AAAGAGGGCC	ACGTGGGAGG	GGTAGGGGTC	GTGTGCAACC	AGGGATCCCA	CCTTCTCTAC	GGTATGTAAA	CACATGTCCC
5810	5820	5830	5840	5850	5860	5870	5880	5890	5900
GGAGGAGGGT	TAGGTTCTTA	CACATAACCGA	ACATTACAT	CCGGTGCACT	GGTCCCGAGC	GGCGGGCCCC	CCATATTTTC	CCCCGCCCTG	AGACAAACAG
CCTCTCCAC	ATCCAAGAA	GTUATTGGCT	TGTAAGTGTA	GGCCACGTGA	CCAGGGGTCC	CCGCCGGGGG	GGTATAAAA	GGGGGGGACC	TCGTGTCGTC
5910	5920	5930	5940	5950	5960	5970	5980	5990	6000
GAGTACACAG	AGCCCTAGCG	ACAGGTCTCT	CGGGTCGACA	ACCCCATCCA	TAGGGGAGAG	CTTACGCCCG	TACTGGAGAC	GTGAGTCCAA	CAGTCAAGAA
F CTCATCTGCT	TCCGGATGCG	TGTCAGGAG	CGCCAGCTGT	TGGGTAGGT	ATTCCCTCTC	GAATGCGGGC	ATGACCTCTG	CATCAGCGTT	GTGAGTTCT
start 1 st leader				splice 1 st leader			ZATA		

FIGURE 26.1 (Continued)

6010 TCCTTGCTCC AGGACGAGG	6020 TCCTAACTA AGGATTGAT	6030 TAACCTGTCAT ATTGACAGTA	6040 GGTCGTCTCT CCAGCAGAGA	6050 ACGGAAAAATA TGCTTTTAT	6060 TTCTGAGAGC AAGACTCTCG	6070 AGGTAGACCA TCCATCTGGT	6080 GTCTTTTGTG CAGAAACAC	6090 TTAGAAAGAC AATCTTCTG	6100 AACAGGTGUA TTGTCAGCT
6110 ACCACCGTTT TGGTGGCAAA	6120 ACTAGCTATC TCATCATAG	6130 TCCCGCAACC AGGGGCTGG	6140 TATCTTCGAA ATAGAAGCTT	6150 CCGCTACCTC GGCGATGGAG	6160 CGGTACCAAA CGCATGGTTT	6170 CCAGAAAGAG GGTCTTTTTC	6180 GGACAGGCGC CCTGTCCGCG	6190 GCGAGGAACC CGCTCCTTGG	6200 GCCGCTACAA CGCGATGTT
6210 TTCGACCTGC AAGCTGGACG	6220 ATGAGCGCGC TACTCGCGCG	6230 GGTGTGTAAA CCACACATTT	6240 CGTAAGTCCCT CCATTTCAGGA	6250 TTCTACCAAC AAGATGGTTG	6260 AGTCAAGTAG TCAGTTGATC	6270 GCCTTGATAA CGGAATCTAT	6280 GACTGAGCGG CTGACTCGCC	6290 TAGGGGATAA ATCCCTATT	6300 CACGTCCCAA GTGCAAGGTT
6310 TAGCTTAGGT ATCAGATCCA	6320 GTGACACCGC CACTGTGGCC	6330 GTGGAGCGGA CACCTCGCCT	6340 GCTTCCCGGA CGGAGGGGCT	6350 GTAACCAAGT CATTTGGTCCA	6360 CGTCTCAGCT GCAGAGTCGA	6370 GGAGGAAAG CCTCTTTTTC	6380 AACTTGTCTT TTGAACAGAA	6390 TCCCTCTCTC AGGGGGGAGG	6400 CCAGATCTGT GGGTCTAGCA
6410 ACTTGGATAG TGAACTCATC	6420 TCCCTCCAGG AGGGGGTCC	6430 CGTAGATACC GCATCATATG	6440 ATTATTAAGG TAAATATTCC	6450 GCCATCGTTT CGGTAGCAAA	6460 AGAAACAGTT TCTTGTCTAA	6470 TTATCGACTA AATAGCTGAT	6480 CCACGCTCCT GGTGGCAGGA	6490 AGTAGGTTCC TCATCCAAAG	6500 ATAGAGGGT TCATCTGCCA
6510 AAGAGCTTGA TTCTGAACT	6520 CGGTCCGCGC GCCAGCGCGC	6530 CGAGTATCCC GCTCATAGGG	6540 CAATTCTCCC GTTAAGAGGG	6550 CAGCGGGTCC GTGCCCCAGG	6560 CGTACCCAC GCATGGGGTG	6570 CCACTCGCGC GGTAGCGCGC	6580 CTCCGTATGT GAGGCATACA	6590 ACGGTGTCTA TGCACACAT	6600 TAGCATCTGT TACGTAGACA
6610 ATCTCCCGCA TAGAGGGGCT	6620 GAAGCTCTTA CTTCGAGAA	6630 CGGCTACATT GCGGTGTAA	6640 CACCTTATTC GTGGGATAAC	6650 TCGCGGGGGG AGGCGCCCTC	6660 AGACTACGAA TCTGTATGCT	6670 CGAGCTGTGA GCTCCGACAT	6680 TCAGTATCTC AGTCTAGAG	6690 AAGTACACTC TTCATGTGAG	6700 CCCCCTCTCT GGGGCGAGGA
6710 CTGGCCCGCG GACCCGGGCG	6720 GCTAAACAC CAGATTGGTG	6730 GCCAACCCAA CGGTGTGGTT	6740 AAAGCGCGGA TTTCCGCCCT	6750 CATTTGCTAA GTTAAACATT	6760 ACCGCTTTCT TGGCGAAGAA	6770 ACCGTACCCT TGGCATGGGA	6780 TAACCTTCTT ATTGGAAGAA	6790 TATCATCCAG ATATGAGTCT	6800 AGACCTTATA TCTGGAATAT
6810 CAATTITTAAT GTTAAATGA	6820 TGACTTCCAT ACATGAGGTA	6830 CGGGATGTCT GGCTACAGGA	6840 CAGAGAAATC GTCTCTTATG	6850 TTACCCCGTA AAGTGGGCTG	6860 TACTGAGAAC ATGACTCTTG	6870 GTGCAACCGA CAGCTTGGCT	6880 TGGTGGAGAC ACCAGCTCTG	6890 GCCACTGTCT CGGTGACAG	6900 ATGTAGTCTC TACTCCAGG
6910 CGGTGATCA GCACAGTGT	6920 ACCTCTCAAG TGAGAGTTTC	6930 GACCTACTAC CTGGATGATG	6940 AGTATTGCGC TCATAACCGC	6950 CAACCGAAAA GTTGGCTTTT	6960 GAAAGGGGTG CTTTTCCCTC	6970 TCGAGCGCCA AGCTCGCGGT	6980 ACTCTTCCAT TGAGAAAGTA	6990 AAGAAGCACT TTCTTCTGTA	7000 AGGAGGTCTA TCTTCCAGT
7010 TGAGAGAGCT ACTTCTCGAG	7020 CCCTTTGGGC GGGAACCCCG	7030 AGAAAGAAC TCTTTTCTG	7040 GTGCAATCTC CAGGTAAGA	7050 CGGGTTGTAC GCCCAACATG	7060 ATCTTGACTA TAGAACTGAT	7070 ACTGACGGA TGACTGCCTT	7080 CATCCCTGTC GTAGGACAG	7090 GTAGGAGAA CATCCCTCT	7100 GGTAGCCCTC CCACTGGGAG
7110 TCTCATAGGA AGAGTATGCT	7120 ACCGGACGTA TGGGCTGCAT	7130 ACCGCTCGCT TGGCGAGCGA	7140 CCATACTCAC GGTATGAGTG	7150 TCCCGTTTTC AAGGCAAAAG	7160 ACAGGGGACTG TGTCCCTGAC	7170 GTACTGAAAC CATGACTTTG	7180 TCCCTTAACTA AGGAATGAT	7190 TGAACCTTCAG ACTTGAATC	7200 CTACAGTAGT GATGTCTATC
7210 GTCCGGGGGA CAGGCCCTCT	7220 CAAGGGTCTC GTTCCCGAG	7230 AACTTTCAG TTGGAGATTC	7240 CGGGCGAAGA GCCCCCTTCT	7250 ACATCCGCC TGTAGGCGGG	7260 TAACCCGTTT ATTGGGCAAA	7270 CGCTTTCATT GCGAAAGTAA	7280 GTAGTAACCT CATCATTTGA	7290 CTCCTAGAGT GAGGATCTCA	7300 GGCCGGGACC CCGGCCCTCG
7310 CGTACTTTAA GCATGAAAT	7320 AGCCCACTAA TCGGGTGATT	7330 AATTTTCCGA TTAAAAGGCT	7340 CTCCCTGGAG GAGGGACCTC	7350 ACGAGCCAAT TGCTCGGTTA	7360 AAGTATGGA TTGATAAECT	7370 CTCCGCCGTT GAGCGGCCAA	7380 CTGCTAGAGT GACCATCTCA	7390 AGTTTCTGTA TCAAAAGCAT	7400 ACTACACAC TGATGTGTTG
7410 GGGGTGATAC CCCCACTATG	7420 ATGTCAAGAT TACAAUUTCTA	7430 TCTTAGCTCC AGAATCGAGG	7440 CCAGCGGGAC GGTGCCCCCT	7450 TGTACTCCGT ACATGAGGCA	7460 CGAAGAATCT GCTTCTTGAG	7470 AAGAAGTTTT TTCTTCAAAA	7480 CACTCTAGAC GTGAGATCTG	7490 ATCCCACTCA TAGGGTCACT	7500 CTCTCGATAC GAGAGCATAG
7510 ACAAGCTCCC TGTTGAGGG	7520 GGGTAAAGAC CCCATTTGCT	7530 GTACACTCCC CATGTGAGGG	7540 AAGCGAAACT TTGCTTTTGA	7550 CCTTCTCTCT GGAAGGAGGA	7560 GGTCTCCAGG CCAGAGGTCC	7570 TGACGGTCC ACTGCCAGTG	7580 GACAACATT CTGTTTGTAA	7590 GACCAAGGCC CTGGTCCCGG	7600 ATGACTGCTT TACTGACGAA
7610 TTACGACAGG AATGCTGTCC	7620 CTGACGGTAG GACTGCCATC	7630 AAAAGACCCC TTTTCTGGGG	7640 ACTACGTTAT TGATGCAATA	7650 CTTCCAAACC GAAGGTTTGG	7660 CCAGGAGCGG GGTCTCTGCC	7670 CGGTCCGTAG GCCAGCGATC	7680 GGTGAACCTA CCACTTGAGT	7690 AAGTACCGCT TTTATGGCGA	7700 ACAGTATCCG TGTATAGGCG
7710 CTACAATTGC GATGTTAAGC	7720 TCGGCGACCA AGCCGCTGGT	7730 GAGGTCTCTC CTCCAGAGAG	7740 AAGTACTGG TTTCATGACC	7750 TCGTACTTCC AGCATGAAGG	7760 CCTAATCGAC GGATTAGCTG	7770 GAAGCGTTTC CTTCCCAAG	7780 CGGGGGTAGG GCCCCCATCC	7790 TCCATATCCA AGGTATAGGT	7800 GAGATGAGC CTCTCATCTG
7810 ATCCACTCTT TAGGTGAGGA	7820 TCTCGGAAG AGAGCTTTTC	7830 ACACGCTCTC TGTGCGAAGA	7840 ACTCTCGGCT TGAGAGCCGA	7850 AGCCCTTCTT TCGGGGAAGAA	7860 GACCTAAAGG CTGGATTTC	7870 ACGGTGGTCA TGCCACCACT	7880 ACCTCTTAC TGGAGGAATG	7890 CGCAACTAC GCTGTGATG	7900 ACTACCTTCA TGATGGAAGT
7910 TCTTGAAGGA H AAGACTCCCT	7920 CGCTGCGCGG GGCAGCGGCG	7930 CTCGTAAGTA GAGCATTCAT	7940 CGAACACGAA GCTTGTGCTT	7950 CATGTCTGCC GTACAGACGG	7960 GGCTCATGTA CCGACAGTCT	7970 GGTGTCTGTA CGCAGGATTT	7980 GTGCTCTACG CAGCGGATGC	7990 TGGAGTACT ACCTCATGAA	8000 ACTCAAGATG TGATGTGATC

FIGURE 26.1 (Continued)

8010 GACTGAAGGA CTGACTTCT	8020 AACTGCTCT TTGACGAGAA	8030 TAAAGTCAAC ATTTCAGTGG	8040 TTTAACTCC AAAATTGAGG	8050 GAACCGCGAA CTTGGCGCTT	8060 CATGGAGCGC GTACCTCGCG	8070 GAGATGATAC CTCTACTATG	8080 AACAGACGTA TTGCTGCTAT	8090 GCCGTACTGG CGGCTAGACC	8100 TAGAAGACAG ATCTCTGTCT
8110 AGCTACCAAC TCGATGGTGG	8120 AGTAGCACTG TCATGCTGAC	8130 CTCGGAGCGG GAGCCCTCGC	8140 CCCTCCG7TC GGGAGGCAAG	8150 AGGTCTGGAG TCCAGACCTC	8160 CCGCGCCGTC GGCGCGGCAG	8170 CCCGCTCGA GGCGGAGCT	8180 GCTCTGCTC CGAGGACGAG	8190 TCGCGCGTCC AGCGCGCAGG	8200 GGCTCTGACA CCGAGCTGT
8210 GGTCCGAGGA CCAGGGTCTT	8220 CTCTGCGAGC GAGACGCTGC	8230 CCTCAGTCCA GGAGTACAGT	8240 ATCATCCGTC TAGTAGGCAAG	8250 ACAGTCTCTT TGTAGAGAGA	8260 AACTGAACGT TGTACTTGCA	8270 ACTAGAAAAG TGATCTTTTC	8280 CTCCCGCACT GAGGGGTGCA	8290 CCCTCCAAAT GGGAGTTTCA	8300 CTACCATGAA GATGTAATT
8310 CTAGAGGTGC GATCTCCACG	8320 CCAGGCAACC GGTCCGTTGG	8330 ACCTCTACAG TGGAGATGCT	8340 CTACCGAACG GATGGCTTGC	8350 TCCCAAGGCA AGGGTTCCGT	8360 CGGGGAACCC GCCCTTTGGG	8370 CGGATGGTGG CGCTACACCC	8380 CAGGGGAACA GTGCGCTTGT	8390 AAAAGGAAAA TTTTCCTTTT	8400 CCCCCGCCCA GGCGCGCGGT
8410 CGGAGCAAAC GGCTCTGTTG	8420 GAGAAAGCTA CTTCTTGCAAT	8430 CAATCTCTCG GTTTAGAAGC	8440 CCACAGCTCC GGGTGTCGAG	8450 CGCCGCTGCG GGCGGACCGG	8460 CGCCGCTCCC GGCGGACGCG	8470 CGCCGAGCCC GGCGCTCGGG	8480 TGGCGCGCGG ATGCGTGGCA	8490 TACCGACCGT ATGCGTGGCA	8500 CACCATGCGA GTGGTAGCTC
8510 CCGTGGCGCG GGCAGCGCGC	8520 CGCCCATCCA CGCGGTAGGT	8530 AGACCATGAC TCTGTACTGT	8540 CGCGGACTCT CGCCCTGAGA	8550 TCGAGCGCTA AGACTCGCAT	8560 CGCGCTGCTG CGCGGACGAC	8570 CGCCGCAAC CGCGCGTTTG	8580 TGAGGACCTT ACATCTCTGA	8590 AGACTGCGGA TCTGACGCTT	8600 GACCACTTTT CTGGGTGAAA
8610 CGATGGCGCG GCTACCGCGC	8620 GGCACTCGAA CCGTGAGCTT	8630 CTTGGACTTT GAACCTGAAA	8640 CTCTCAAGTT GAGAGTTCAA	8650 GCTTTAGTTA CAGAATCAAT	8660 GAGCGATAGC CTCGGTATCG	8670 AACTGCGCCC TTGAGCGCGG	8680 GAACGATATC CTTGGCTAGG	8690 CTAAAGAACG GATTTCTTGC	8700 TGCAGCGGTC ACGTGCGCAG
8710 TCAACAGGAC AGTTGTCTCT	8720 CATCCGCTAG GTAGCGGATC	8730 AGCCGGTACT TCGCGCATGA	8740 TGACGAGCTA ACTGCTGATC	8750 GAGAAGGAGA CTCTTCTCTT	8760 ACCTCTAGAG TGGAGATCTC	8770 CGCGCGCGCG CGCGCGCGCG	8780 AGAGAGCTGC TCTCTGAGCG	8790 CACCGGCGCT GTGGCGCGCA	8800 CCAGCAACCT GGTGTAGTTA
8810 CTACCGCGGT GATGCGCCCA	8820 TACTCAACTC ATGAGTTGAG	8830 TCITTCTGTA AGAAAGCAAT	8840 GTACGGCGCG CATGCCCGCC	8850 AGCAAGGCTT TCGTTCCAGA	8860 GGCGGACAT CGCGCTGCTA	8870 CTGGTGTGCG GACCCACGCG	8880 GGGTGCGCTA CCCACGGGAT	8890 GAGAGCGCGC CTCTGCGCGC	8900 GTACTGCTGG CATGACCAAC
8910 ACCCGCTCCA TGGCGAGGTT	8920 ACTCGAGGTT TGAGCTCCAC	8930 CACCGCCGAC GTGCGGGTGG	8940 TTCTGGCGTA AAGACCGCAT	8950 TCAACGTATC AGTTGCAATG	8960 CGCGACCTTT CGCTTGGAAA	8970 TCCATCAACT AGGTAGTTGA	8980 CACACCAACG GTGTGGTGGC	8990 CTACACGAGC GATGTGCTCG	9000 CACTGCTTCT GTGACGAAGA
9010 TTATGTACTA AATACATGAT	9020 GGTAGAGAGG CCATGCTCTC	9030 TCGCGGTAGA AGCGGCATCT	9040 CGCACTGTAG CGCTGACATC	9050 CGGGTCGCGA GCCGAGCGCT	9060 AGGTTCCGCA TCCAAGCGCT	9070 GGTAGCGGAG CCATGGCCCT	9080 CATCTTCAAG GTAGAAATTC	9090 TGCGGTITCA ACGGCAAGTT	9100 ACTTTTTCAC TGAAAAAGTT
9110 CTCTAATGCC GGAGTTACGC	9120 CGCGTGTGCC CGCGACACGG	9130 AGTTGAGGAG TCAACTCTCT	9140 AAGGTCTCTT TTCCAGAGAA	9150 GCCTATTCAA CGGATTAAGT	9160 GCCGCTACCA CGCGGATGCT	9170 CCACGCGTGG GGTGGCGACC	9180 ACCGCGAGCT TCGCGCTCGA	9190 TTGAGGATC AAGTCTCTAG	9200 CTAAGAGAGG GATTCTCTCC
9210 AGTTAGAGAA TCAATCTCTT	9220 GAAGAAGGTT CTTCTTCCAC	9230 ATTGTAGAGA TAACATCTCT	9240 AGGAGAAATC TCTCTTCCAG	9250 CACCCCGACG GTGGGGCTGC	9260 TCCCTCTCCC AGGAGAGGCG	9270 CCTTGGCGCG GGAAACGCGC	9280 CTGCGCGCGC GACGCGCGCG	9290 CGCGTGGCGG CGCGACGGCG	9300 TCTGCGAGCT AGACGGTCTGA
9310 ACTTAGAAGA TGAATCTTTC	9320 TTACTGGAGA AATGACTCTT	9330 GGCGCGCGCG CGCGCGCGCG	9340 CGCGGTATAG GGCGCATAGT	9350 GAGGCACTGC CTCGGTGAGC	9360 CGTGTGCGCA GCACGACCGT	9370 AGAGGAGACC TCTGCTGCGG	9380 AGAGTCTCAC TCTGAGAGTG	9390 TCTGCGGAGG AAGACGCTCT	9400 CGCGGTAGAG CGCGCATCTC
9410 GGACTTCAAC CCTGAAGTGG	9420 ACTGACCCCT TGACTGGGAG	9430 CGAGAGGCAA GGTCTCCGTT	9440 CCCGTCCCTG GGGACGGGAC	9450 TGCGCGGACT ACCGCGCTGA	9460 AATACGTAAA TTATGCAATT	9470 ATAGTTAAGC TATCAATTGC	9480 GGGCATCCAT CCCGTAGGTA	9490 GAGGCGCGGT CTCGCGGCAA	9500 CCTGGACTAA GGACCTGATT
9510 CAGAGTTCTA GTCTCAAGAT	9520 GCTGCGCTAG CCACGGGATC	9530 ACTTTTGGAA TGAAAACCTT	9540 AGCTGCTTTC TCGACGAAAG	9550 GCAGATTGGT CGTCTAAECA	9560 CAGCGTATAG GTGCGAATCG	9570 GTTCATCTCC CAAGGTAGGC	9580 ACTGCTGACA TGAGCACTGT	9590 AAGAACGCCC TTCCTGCGGG	9600 GCCCCCGCGC CGGGGGCGCG
9610 ATCTGCGAGC TAGAGCTCTC	9620 CAGCCCAAGG GTGCGGGTTC	9630 AGAGAAAGAA TCTCTTTTCT	9640 GAGGAAGGGG CTCTCTGCCC	9650 GAGAACGCTC CTCTTGGCAG	9660 CCACTCTGCT GGTGAGACGA	9670 ACGACGACCA TGCTGCTGGT	9680 CTACTTTAAT GATGAATATA	9690 TTTATCCGTC AAATAGGCGA	9700 AAAAATCTGC TTTAAAGACG
9710 CGCCTACCAAC GGGATGGTGG	9720 CGCTCTCTGT CGGAGGAGCA	9730 GGTTCAAGAA CCAAGTCTTT	9740 CCCAGGCCGA GGGTCCGAGT	9750 ACAACCTACG TGTGAGATGC	9760 CGTCCGCTAC GCAGGCGATG	9770 TCGGTAGGGG AGCCATCCCC	9780 GTTCTAGTGA CAAGCATCAT	9790 GGACTGTAGA CCTGACATCT	9800 CCGGTCTAGA GGCCAGATCT
9810 AATATCTCTA TTATATAGAT	9820 GAACGTACTC CTTGATGAGG	9830 AGCAAGGTGC TGTGTTCCAG	9840 CCGTGAAGAA GGCACTTCTT	9850 GAAGCGGGCG CTTGGCGCGG	9860 GGACGGTACG CTTGGCATGC	9870 TAGCTCTACT ATGCGAGTGA	9880 AGGGCTTGGG TCCCGAACCC	9890 CGCGTACCCG GGCATGGGCG	9900 ACCTGTTCAC TGGACAAAGT
9910 GGTGGAGGCG CCAGCTCTCG	9920 ATGTTGGGAA TCAACCTCTT	9930 AGCCGCTCTT TCGGCGAGGA	9940 ACCGAACGAC TGGCTTGTCT	9950 GTGGACCCAC CACCCTGGGT	9960 TCCGACCGAA AGGGTGGGCT	9970 CTTTCAGCAG GAAAGTCTCT	9980 TTTCAGGTGC AAAGTCCACG	9990 TTCGCCACCA AAGCGGTGCT	10000 TCCGGGGCCA AGGCCCGGCT

FIGURE 26.1 (Continued)

10010 CAACTAACAC GTTGATTGTG	10020 ATCCTCGTCA TAGGAGCAGT	10030 ACCGGTACTG TGGCCATGAC	10040 ACTGGTCAAC TGACCACTTG	10050 TGACAGACCA ACTGTCGTGT	10060 CGGGTCCCCG GCCCAGGGGG	10070 GTGCTCGAGC CACGAGCTCG	10080 CACATGAAGT GTGTACTTCA	10090 CCGCGCTCAT GGCGCGAGTA	10100 ACGCGCGCAC TGCGCGGTG
10110 AGTTTCTACA TCAAAGATGT	10120 TACCAACGT AATCCTTCCA	10130 CCACGCGTGG GGTGCACACC	10140 TCCATGACCA AGGTACTGGT	10150 TGGGCTACTC AGCCGATGAG	10160 TTTACACCGG AAAGTGTGGC	10170 CTACCGACCG GATGGCTGGC	10180 CCATGTCCCC GGTACAGGGG	10190 GGTAGCGAGA CCATCGCTCT	10200 CATCGGCCCC GTAGCCGGGG
10210 GAGGCCCGCG CTCCGGGGGG	10220 CTCCAGAAGG GAGGTCTTCC	10230 TCGTACTCCG AGCATGAGGC	10240 CCACCATCGG GGTGGTAGCC	10250 CATCTACATG GTAGATGTAC	10260 GACCTGTAGG CTGGACATCC	10270 TCCACTATGG AGGTGATACC	10280 CCTCCGCCAC GGAGGCGGTG	10290 CACTTACGTG GTGGATGCAC	10300 CACCTCTGAG GTGGGAACTC
10310 CGGCTGCGCC GCGACGCGGG	10320 AAGGCTACCA TTCCAGATGT	10330 ACCGGTGCGC TGCGCAGCGG	10340 GTACTTGATC CATGAAGTAG	10350 AAGTACCATC TTGATGGTAG	10360 CGTGCCAGAC GCACGGTCTG	10370 CGGTCACTCC GCCAGTGAGG	10380 GGCGCGGTCA CGCGGCGAGT	10390 GTAACTCGGA CATTTGACGT	10400 GACATCTGTG CTGTAGACAC
10410 CCTCTTTTTC GGAGAAACGG	10420 TTTCGCTACT AAAGCGATGA	10430 CGCCGAGCTG GCGGCTCGAC	10440 AGGCACCGGA TCCGTGGCCT	10450 CCCCCTTTGA GGGGGAACGT	10460 CCTGCCCAAC GGACGGGTTG	10470 CCAGCGCCAC GGTCGCGGTG	10480 ATGGGCGCAA TACCCCGGTT	10490 GCTCAGGTTT CGAGTCCAAA	10500 CGATTCTGTA GCTAAGCAAT
10510 GTGTAGGCTT CACACTCGGA	10520 AGCGGCGCTC TCGGCCGGAG	10530 GGCGCGGATT CCGGGCTATA	10540 GCACCATTAAC CGTGGTATTG	10550 CGATAGGGCA GCTATCCCGT	10560 GAGCTGGGTC CTCGACCCAG	10570 GGCTGCTTAT CCGACGAATA	10580 AGGTCCCATG TCCAGGGTAC	10590 CCTCATCTCA GGAGTAGAGT	10600 GCAAAAACGA CGTTTTTGCT
10610 CGAAAAAGGG GCTTTTTTTC	10620 ACCTGCACAC TGGAGGTGTG	10630 GGTAACGGTG CCATTGGCAC	10640 CAGTTGAAAA GTCAAGCTTT	10650 TGTTCGAGGT ACAACGCTCA	10660 CAAGAGCCCC GTTCTCGGGC	10670 GCACTCACCG CGTGAGTGCC	10680 ACGCGGGGCA TGGCGGCGGT	10690 TCAGACCTCT AGTCTGGAGA	10700 TACTAGCGGG ATCAGTGCCT
10710 TCCCAAGCCA AGGGTTGCGT	10720 ACGCCATAGC TGCGGTATGC	10730 GGGGCCCAAGC CCCGGTTTCG	10740 TCGGATTGCG AGCCTAAGCG	10750 GCGGAGCATA CGGCTCGTAT	10760 GCGGGCCAAA CGGCGGTTT	10770 GGCGCTGTTT CGGCGACAA	10780 GCTCCCATAC CGAGGGTATG	10790 CGTCGGGTCA GCAGCCCAAT	10800 GTAAAGGTTT CATTTCGAAG
10810 TGGGGCGGTC ACCCGCGCAG	10820 GGCTGAAGAG CCGACTTCTC	10830 GTCAATGCTC CAGTTTAAAG	10840 CTCGCTCGGG GAGCGAGCCC	10850 AAAAAAAAAA TTTTTTTTTT	10860 AACAAAAAGA TTGTTTTTCT	10870 GCGGGTCTAC GCGCCAGATG	10880 GTAGGTACCG CATCCAGTGC	10890 ACGCTGTCTA TGGACAGAT	10900 CGCGGGGCTC GCGCCCCCAG
10910 K GTGTCCGGG CAACAGCGCC	10920 GAAGAGTGGT CTTTCAGCA	10930 TGTGGTGGTT ACAGCCACAA	10940 TTCCGAGAAG AAGGCTCTTC	10950 AACGAGGA TTGCTCT	start 52, 55 kD proteins				

FIGURE 26.1 (Continued)

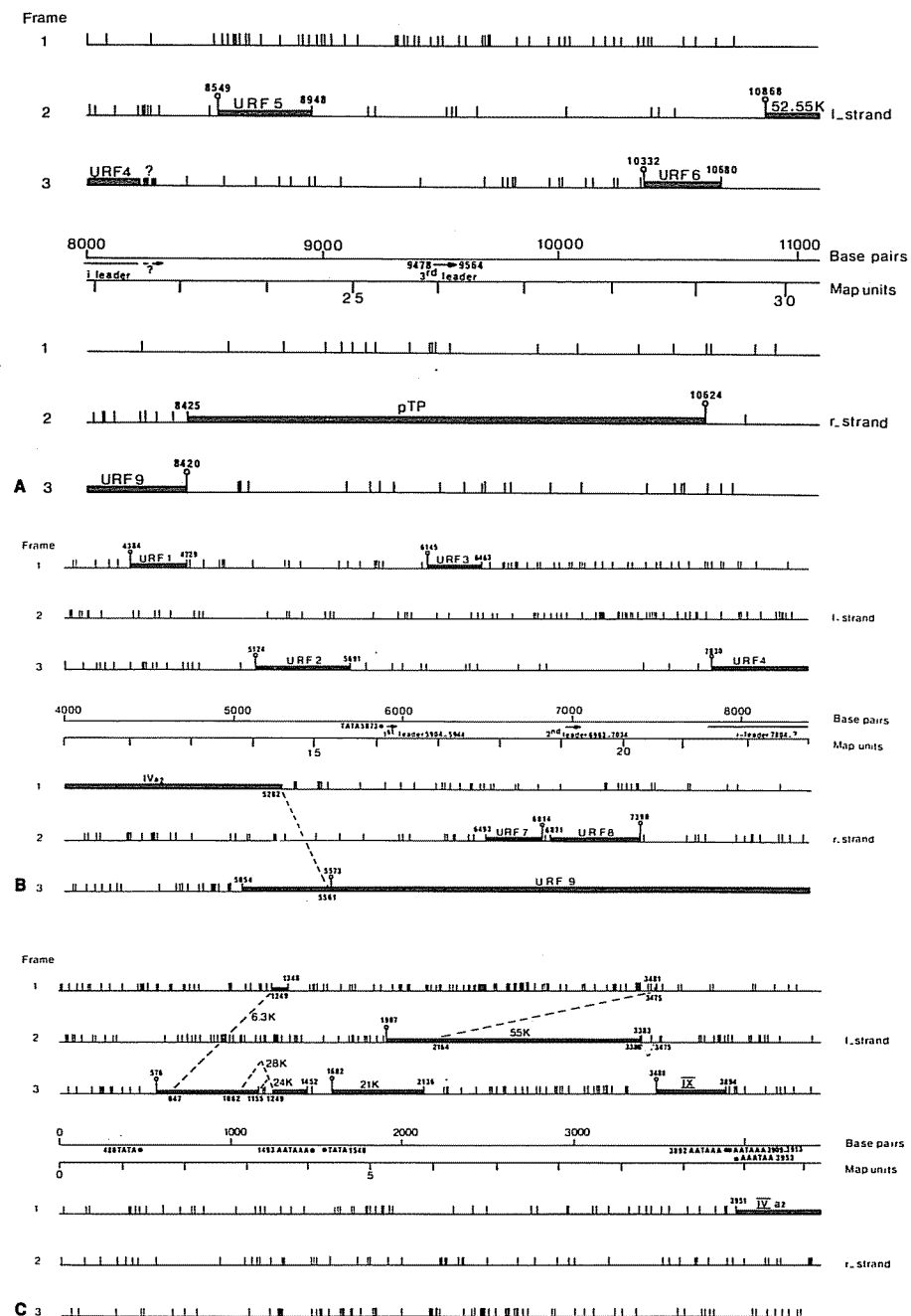


FIGURE 26.2A-C. Structural organization of a region between coordinates 0 and 31.7 on the Ad7 genome. This map is derived from the nucleotide sequence in Fig. 26.1. For details, see Fig. 3 (Section VII).

10	20	30	40	50	60	70	80	90	100
CTATATATATA	TTATATGGAA	TATGACCTGA	TCACGGTAT	AATTTTACTT	CACCGGCATC	ACACATTAAA	CTAACCCACC	TCACACCCGA	AACCGCACGA
CTATATATATA	AAATATACCTT	ATACTGGACT	AGTCCCAATA	TTAAATGAA	GTGGCGGTAG	TGTGTAAATT	GATTGGGTGG	AGGTGTGGCT	TTGGCGTGGCT
110	120	130	140	150	160	170	180	190	200
ACATTCAAC	CGGCTACTC	CTTACCCCG	CGCCGACCC	TCGGCCCGCG	CGGCTACAC	TGCAAAATCT	CGCGTAAAT	GTGCTTTAC	TACAAAAAC
TGTAAGTTTG	GGCGGATGAG	GAAGTGGGG	GGGGCGTGGG	AGCGGGCGCG	CGCGGATGTG	ACGTTTATGA	CGCCATTTTA	CACGGAATG	ATGTTTTTTG
210	220	230	240	250	260	270	280	290	300
CGGCAACAA	CACGTTTAAA	ACACAAAATC	CGCGTTTITG	ACTTTACGCC	TTCACTTTTA	ACTACTGCGC	TTAAATAAT	ATCCGCGCCT	TATAAATGGC
GGCGTTGTTT	GTGCAAAATT	TGTTTTTATG	CGCGGAATAC	TGAAATGCGG	AAGTGAAAT	TGATGACGCC	AATTTTATTA	TAGCGCGCGA	ATATTATACG
310	320	330	340	350	360	370	380	390	400
TCCGCTCTCA	CTTGAGACTC	GGAGATGCAC	ACCCAAAGCT	ATGCACTGCG	TGCCCCTTTG	AGGTGCAACG	CGAOTTTCCC	CGCCAAATAA	CAAGACAAGT
AGGGCAGAGT	GAACCTCTGAG	CCCTCTAGTG	TGGGTTTCTG	TACBTGAGCG	ACGGGGAAC	TCCAGTTTGC	CGTCAAAAGG	CGGTTTTATT	GTTCCTGTCA
410	420	430	440	450	460	470	480	490	500
GACTAGCAAA	CCCATAAATT	ACGGCGGCAC	AAGCAGTTCT	CGGGTGAGAA	CTCAGGTGCG	CTCTCTCTCA	AAGAGACGGT	CGAGTAAAG	TGCCCGCGTA
CTGATCGTTT	GGGTATTTAA	TGCCCGCGTG	TTGCTCAAGA	GGCCACTCTT	GAGTGCCAGC	GAGAAAGATT	TCTCTGTCCA	GCTCATTTTC	ACGGCGCCAT
510	520	530	540	550	560	570	580	590	600
ATACTCTTGA	CTTACTCTGAG	GGACCAAGGA	CACCATAGTCT	CTTGGACTGC	TGTATAAAGT	CGTAAACCCG	CTGTTGAAAA	AATTGCTCCA	TGGGTCACTA
TATGAGAACT	GAATGACTCT	CGTGTGCTCT	GTGCTATCAG	GAAGCTGACG	ACATATTGGA	GCATTGTGGT	GACAACATTTT	TTAACGAGGT	ACCCAGTGAT
610	620	630	640	650	660	670	680	690	700
CTACTAGAAA	TACAAGGCAG	AGAAATGCTT	GACATACCTAG	AACATACACTT	CAGACGGCCA	CTTCTATTAT	TACTGTCCCG	CCACTTTACTC	AAAAAGGCGT
GATGATCTTT	ATGTTCCGTC	TCCTTACGAA	CTGTATGATC	TGTATGTGGA	GTCTGCCGCT	GAGGATATAA	ATGACAGGCG	GGTGAATGAG	TTTTTTCCCG
710	720	730	740	750	760	770	780	790	800
TTAGCGATA	AAATCGACCG	TCACCTCCCA	ACAAAATGG	CTCCGGAGGA	CATGAAGAG	GACAGACACT	CGGTAACCC	CCGCTTACAT	ACGGGTGTGA
AATCGCTTAT	TTTAGCTGCC	AGTGAGGGGT	TGTTTTTACC	GGAGCCCTCT	GTACTTTCTC	TGCTCTGTGA	GCCTATTGGG	GGCGAATGAG	GGGCAACTCT
810	820	830	840	850	860	870	880	890	900
CGTGGGACTT	CTATACCTAA	ATAACACGAT	CGCTACCCCG	AAAGGGACAT	CGCTAAGCCT	TCGCTCTGCT	CTGCTCTTCC	CTTACCGCGT	ACAAAGCGGT
GCACCTTGAA	GATATGGATT	TATTTGCTTA	CGAGATGGCG	TTTCCCTGTA	CGGATTCCGA	AGACGAGCAA	GACGAGAACG	GAATGCGCGA	TGTTTCTGCA
910	920	930	940	950	960	970	980	990	1000
AGGCGCTGAC	GACGACGGCG	ACTATCCCTT	GCACCTCTCA	AAGTCAATCT	GGTAGGTCTC	AACGGGCGCTG	TGTTAACATT	CAGGACACTC	GTGGTGGGCT
TCCGCGCTGG	CTGCTGGCGC	TGATAGGGAA	CGTAGGAGCT	TTACGTTAGA	CCATCCAGAG	TTGCGCGGAC	ACAATTTGTA	GTCTCTGTAG	CACCACTGGA
splice 26 kD E1A RNA									
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
TATCATGACC	TTTATGACTG	AATTACACGA	GAACACGAT	ACACACTGCG	ATGTTGTATCA	AGTAAATGTC	ATTACACAGA	TACCTCCACG	CTTCCACTAA
ATAGTACTGG	AAATAGCTAC	TTAATGTGCT	CTTTGTGCTA	TCTCGAGCC	TACAACATGT	TCATTTACAG	TAGTGTGGCT	ATGGGAGGTT	GGAGGTGATT
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
AAAAAAGAA	TTGCTCACTT	TTTATATATA	AAACAACAAA	ATCCAGGACA	AAGGCTATTA	CTCGGACTTG	GATTATCGTG	AAACCTACCG	CTACTCGCTG
TTTCTTCTT	AAGCAGTGAA	AAATAATATT	TGTTGTGTTT	TAGTCTCTGT	TTCGATAAT	GAGGCTGAAC	CTAATAGCAG	TTTGGATGGC	GATTAGCGAC
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
GGATGGGGG	CTTTGATGCT	TCACGCAAG	GTCTTCTCTA	TTATTTTGGG	CACGGAGTGG	CCCACTGACC	CTCCGCATCT	ACACGACACC	TTTCGTAAAA
CCTCACCCCG	GAACATAGGA	AGTGCGGTTG	CAGAAGGAGT	AATAAAACCT	GTGCTCTCAG	GGGTGACTGG	GAGGCGTAGA	TGTGCTGTGG	AAAGCATTTT
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
CCTAAACTAA	GTCTCTCTTC	TTTCTCTTGT	TTGTCAACGA	CAACTAGACA	GTCACTTTGG	GGGATCTAGA	TTAATTTACCT	GAACACTGCT	GACCCGTAT
GGATTTGATT	CAAGAGGAAG	AAAGAGAACA	AACAGTGCTT	GTGTATCTGT	CAGTGAAACG	CCCTAGATGT	AATTAATGGA	CTTTGAGGAC	CTGGGCAATA
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
TTTATCCCCA	TTACACCAAA	AACACTCAGT	ACATATATTAT	TTGACCAAAG	CCAACCTCAC	AGAACAATTA	CAACACAAAC	CGACCAAAAT	TGTCCTTATA
AAATAGGGGT	AATGTGGTTT	TTGTGAGTCA	TGTATAATAA	AACCTGGTTT	GGTTGAAGTG	TCTTTGTTAT	GTTTGTGTTG	GCCTGGTTAA	ACAGGGATAT
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
TTTGACACCA	ACCACACAGA	AACCTTACAA	GTAGAAATCAT	TACCTCAACC	TTTGACACGA	CGTTTCAAAA	GTCTCCCAAG	CGGTGAGGAA	CGTCATATGG
AAAGCTGGGT	TGGTGTGCTT	TTGAATAGTT	CATCTTAGTA	ATGGAGTTGG	AAACTGTGCT	GCAAGTTT	CAGAGGCTTC	GGCAGCTCTT	GCAGTATACC
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
AGATTITTTGT	GAAGTCCAAA	AACCTCCATA	GACAAACGGA	GATGGAATTC	GTTCACCAT	TTATCCCACT	TCTCTCTGAT	ATCTCTCTCT	AAACTTTTGT
CTTAAANACA	CTTCAAGTTT	TTGAGGTTAT	CTGTTTGGCT	CTACCTTAAG	CAAGGTGGTA	AATAGGGTGA	AAGAAGACTA	TAGAGAGGAA	TTTGTAAACA
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
ATAACGGGCT	GACAGGTCCC	GAACACCGAA	GTGATCTGGA	AACAATGGTG	AACCAACAGG	TCCTTTTTCG	CCAGCTAGCG	ATCTTAAAAA	GTAGACACCC
TATTGCGCGA	CTGTCCAGGG	CTTTTGGCTT	CACATAGACCT	TTGTTACAC	TGTTGTTTTC	AGCAAAAGCT	GGTCAGATCC	TTAGATTTT	CATCTGTGGG
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
TGCTTTGCCAA	CGAAGATAAC	GAATAACCG	TTGGTATAAC	CTATTTACCT	CGCTCTTTAG	GGTGGACTCA	ACCTTAATGT	ACGACCTAAT	GTACAGTTAC
ACGACGGTT	GCTTCTATTG	CAACATATTG	GATAAATGGA	GATGAGAAATC	CGGAGAAATC	CCACCTGAGT	TGGGATTACA	TGCTGGATTA	CATGCTCAATG
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
GTGACACCT	CCCGTACCGA	CTTCTCTCC	CAACGTAATA	TGAGCGACCG	CGCCGGAAGC	TGGTACGGCG	CGGACGGCTG	CAACCTTCTC	CTCTCTCTCC
CAGCTGTGGA	GGGCTAGGCT	GAGAGGGAGG	GTTTGCTATT	ACTCGCTGCG	CGCGCTTTTG	ACCATGCCGC	CGGTGCCGAC	GTGCAAGAG	GAGAGGAGG

FIGURE 27.1A–D. Nucleotide sequence of a region between coordinates 0.0 and 11.5 on the Ad12 genome. This sequence and strategic signals were established by Fujinaga *et al.* (1979), Sugisaka *et al.* (1980), Kimura *et al.* (1981), and Bos *et al.* (1981).

2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
TCCTGGCTT	GGGACGGAC	CACCTCTCA	TTTGTACTT	GTGTGCAAG	TTCTTCCGT	ACATGACTG	AGACGGCTC	CCGGATCAAC	GGCTCTACTA
AGGAGCGAA	CCCTGCGTG	GTGGAGAAT	AAACATGGA	CAACAGGTGC	AAGAAGGCEA	TGTACTTGAC	TCTGGCGAAG	GGCTTAGTTG	CGCATGTAT
		<i>stop 19 kD E1B protein</i>			<i>splice 19 kD E1B RNA</i>				
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
TCCTATTTCG	TCCTTTTTTT	TCCTTCAAT	TTCTTTCGAC	GACAAGAATC	ATCCGATTGA	CAATTAGACT	ACAGGGCGGG	CGCAACCTTT	TGACATATAA
AGAGATAGC	AGGAAAAAA	AGAAAGTTTA	AAGGAAGCTG	CTGTCTTAG	TAGGCTAACT	GTTAATCTGA	TGTCCCGCCC	CGCTTTGGAA	ACTGTATATT
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
CCGTCTCTCA	CGTCTACTT	AAAGTCGCG	CACATACCT	AAATGTCTAG	TTTATGTCAA	AACTGTGTTA	TTTTTGGGTG	ACCAATCTCG	GTACCCCTCT
GGCAGGAGT	GCAGGATGAA	TTTCAGCGGG	GTGATATGCA	TTTACAGTAC	AAATACAGTT	TTGAACAATT	AAAAACCCAC	TGGTTAGAGC	CATGGGAGGA
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
ATACCTCACA	CGATAATTTC	GAACAGGATT	TAACCGGAAT	GCAGGACTAA	CATCGATGTC	TTAATGATTT	TGTCTATTGGT	AATGAAGTAC	GGCGATATAA
TATGGAGTG	GCTATTAAAG	CTTTTGTCAA	ATTGECCTTA	CGTCTGATT	GTAGCTACAG	AATTACTAAA	ACAGTAACCA	TTACTTCATG	CGCTATATT
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
TATCCATTGC	CCCGTTATCA	ACTCCATCTA	TGTTGCTGT	CTCAACGAAA	ATCTACAGCT	TACGTCCCAT	ACCCGGGTCC	CCACCCACCA	AACCTACTTT
ATAGGTAAAC	GGGCAATAGT	TGAGGTAGAT	ACAAGCGACA	GAGTTGCTTT	TAGATGTGCA	ATGCAGGGTA	TGGCGCCAGG	GGTGGTGGGT	TTGGATGGAA
2510	2520	2530	2540	2550	2560	2570	2580	2590	2600
AATGTAAATA	TTTACAATCC	AAACGACCTC	TATTCAATTT	TCCGTAATAC	AAGCTTCGAT	TATGGACAGA	ACAGAACGTA	CCACAAATGA	AAGATTGAA
TTACATTAT	AAATGTAGGG	TTTCTGGAG	ATAAGTTTAA	AGGCATTATG	TTCGAAGCTA	ATACCTGTCT	TGTCTTGCTAT	GGTGTTTACT	TCTTTAATCT
2610	2620	2630	2640	2650	2660	2670	2680	2690	2700
ATCATTTGTA	ACACATCTCA	GAACCTTATT	CCAAAGACGA	TCCCGGACAT	GAAAAATACE	TACAACCTTC	CCAAACCAACC	CACTGTGTTT	TTCAATTGAC
TAGTAACATT	TGTGTAGAGT	CTTGGAAATA	GGTTTCTGCT	AGGGGCTGTA	CTTTTATGCG	ATGTGGGAAG	GGTTTGGTGG	GTAGACCAAA	AAGTAACACT
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
AGACATTTT	TCACAAACAA	ACTTTTACCA	CATGAAGCAA	ATTAACATCT	CCCCCTACGT	GTATAATCCG	TATTACGTCG	AAGTCTTTTA	CGGACAAAC
TCGTAAATA	AGTGTGTGTT	TGAAATAAGT	GTACTTCTCT	TAAATGTAGA	GGGGGATGCA	CATATTAGGC	ATAATGACGC	TTACAAAAAT	GGCTGTTTTG
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
ATAATAACTT	CCCTTACCGA	TAAATTTTCG	TATTATACCA	AAACACCCAC	AGACTAGTTT	GATACGCTGC	AAAACAATGG	ACACGACTAC	CTTTAAGCAT
TATTATTGAA	GGGAATGGCT	ATTTTAAAGC	ATAATATGCT	TGTGGGGGTG	TCTGATCAAA	CTATGCGGAC	TTTTTGTACC	TGTGCTGATG	GAATTTAGCA
2910	2920	2930	2940	2950	2960	2970	2980	2990	3000
ATGGAAATTT	TGACAAAGTA	AACACTCTGA	GTACTCTGTA	ACAAACCCAC	ATACACCAAG	ATTGTGCAAA	TACGGGACAT	GGTATGTAAA	TCCGAATTCG
C TACCTAAAA	ACTGTTCTAT	TTGTAGGCCA	CAGTAGACAT	TGTTGGCTTG	TATGTAGTCA	TAACTGTGTT	ATGGCGTATA	CCATACATTT	AGCCTTAAGG
3010	3020	3030	3040	3050	3060	3070	3080	3090	3100
GGCCCATACA	AATCTGGAG	GGTTACATTG	AAGTCGGTGA	GTTTGTAATA	CGACCTTGGG	CTTCACAAAA	GATCTCACAC	AAATTTACCC	CATAAATACT
CGGGGTATGT	TTAGACCTTC	CCAATGTAAAC	TTACGCCACT	CAAACTATAT	CGTGGAACTT	GAAGTGTGTT	CTAGAGTGTG	TTTAAATGGG	GTATTTGATT
3110	3120	3130	3140	3150	3160	3170	3180	3190	3200
ATAGACACCT	TAAATACATC	CAATATTCTA	TATTACTACT	ATGAGCTGTA	GCACACGGCT	TCACACTCAC	ACCATGCTCA	GTAGATCTTG	AAGCAGGGTA
TATCTGTGGA	ATTATGTAA	GGTTATAAGAT	ATAATGATGA	TACTCGACAT	CGTGGCCGAC	AGTGTGAGTG	TGGTAGCAGT	CATCTAGAAC	TTCTGCCCAT
3210	3220	3230	3240	3250	3260	3270	3280	3290	3300
ACACGATTTA	CATTGACTCC	TGCGACTCTC	ACTGGTGGAA	TGGGACAGAA	CGGACGGCTG	ACTGATACTC	AGTTCTACTAC	TTCTGCTGTT	GACTCCATTC
TGTCTAAAT	GTAACTEAGG	AGCTGAGAG	TGACCACCTT	ACCCGTCTCT	GGCTGGGAGC	TGACTATGAG	TCAAGTGATG	ARGACGAA	CTGAGTAG
3310	3320	3330	3340	3350	3360	3370	3380	3390	3400
ACCCACCTCG	ATCCACCCCTA	ATATTTCCTG	ACCTTTCGCA	GATTTTAAAC	AAAAACAAGA	AAATTGTGCT	GCTACTTGCC	TTGATGAGTC	TGTTGCGGAC
TGGGTGGAGC	TAGGTGGGAT	TATATAAGGC	TGGAAGTCAA	CTAAAAATTG	TTTTTGTCTT	TTTAACAGCA	CGATGAACGG	AACTACTCAG	AACAACGGCT
3410	3420	3430	3440	3450	3460	3470	3480	3490	3500
CGCAAAACT	ACCTCCCAAC	AAATCGGGA	TAACTGGAAG	GTCCAATGGT	ATAACCCGGC	CTCATGGCACT	CTTACCAAT	CCTAGATCTC	ACCTGCCAGC
CGCTTTTGA	TGGAGGGGTT	TTTAGCCCTT	ATTGTACTTC	CAGGTTACCA	TATTGGGCGG	GAGTACGTCA	GAATGTGGTA	GGATCTACAG	TGGACGGTGC
3510	3520	3530	3540	3550	3560	3570	3580	3590	3600
TGGACACCGT	GGACGTTTAA	GTAGTTGTAA	TTGGATACGT	TGATAACCTG	GGAGCGGAAA	CCATGGCGGG	CGGCGGCGAC	GTGGAAGGCG	CGGCGAAGA
ACCTGTGGCA	CCTGCAAAAT	CATCAACATT	AACCTATGCA	ACTATTGGAC	CCTCGCCTTT	GGATAACGGC	GGCCGCGCTG	CAGCTTCCGC	GGCGGCTTCT
3610	3620	3630	3640	3650	3660	3670	3680	3690	3700
TGCGGAGCGT	CATACGGTCC	ACTAAGCTCG	AAGATGTTAG	TGAACCGAAG	CTTACGACAC	TGTGCGTGGC	GTCAAGCTCT	CCTGTAAGAC	TGACAAATAC
ACGGCTCGCA	GTATGGGAGC	TGATTTTCAGC	TTCTACAATC	ACTTGGCTTC	GAATGCTGTG	ACACGCAACG	CAGTTCCGAGA	GGACATTCTG	ACTGTATTGC
3710	3720	3730	3740	3750	3760	3770	3780	3790	3800
AACGGTTCGA	ACTTTGAGAT	TGACGAGTGC	ACCTTCTCGA	TAGCGTCTTC	CAACTCCCTTA	ATCGACTACG	ATGATGGGTA	TGGGTCGGG	TTGACATTG
TGCGCAAGCT	TGAAACTCTA	ACTGCTCAGC	TGGAGAGGCT	ATCGCAAAAG	GTTGAGGAAT	TAGCTGATGC	TACTACCACT	ACCCAGCGCC	AACCTGTAA
3810	3820	3830	3840	3850	3860	3870	3880	3890	3900
GGTTATTCTT	TTTTTGAATT	TAACTCTACC	ACAATACTTA	GAATAAATTA	TGAACAAAAA	AGACTGTACC	ATTCGAGAAC	TGGTGGCAAG	GGATAGTAA
CCAAATAGGA	AAAAAATTTA	ATTGAGATGG	TGTTTGAAT	CTTTATTGAT	ACTTGTGTTT	TCGACATGG	TAAAGTCTTG	ACCACCGTTC	CCTATCATTA
<i>stop protein IX</i>									
3910	3920	3930	3940	3950	3960	3970	3980	3990	4000
TCTGTGTCCA	CTTACACAA	GTCAATAAAC	ATTCTTACTC	GGACATATAA	TTCATG				
D AGAACCGGT	GAATGTGTT	CAGTATTTTG	TAAAGATGAG	CTGTATATT	AAGGTAC				

FIGURE 27.1 (Continued)

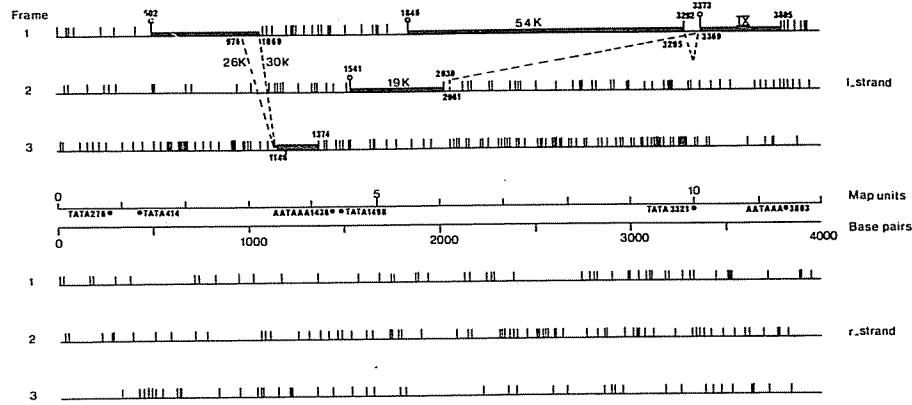


FIGURE 27.2. Structural organization of a region between coordinates 0.0 and 11.5 on the Ad12 genome. This map is derived from the nucleotide sequence in Fig. 27.1.

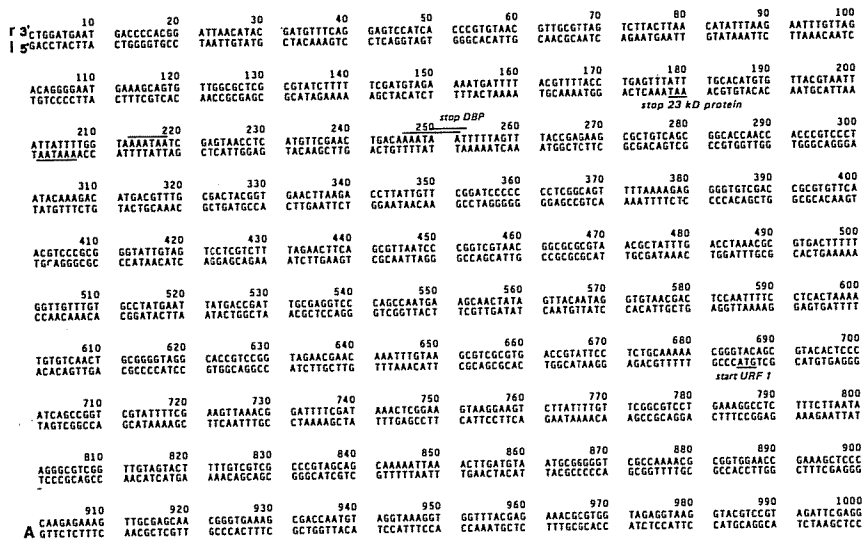


FIGURE 28.1A, B. Nucleotide sequence of a region between coordinates 61.5 and 67.0 on the Ad12 genome. This sequence was established by Kruijer *et al.* (1983).

1010 GGAAGCGCGA CCTTCGGCT	1020 GCCATGTGAA CGGTACACTT	1030 TACGAGGGTG ATGCTCCAC	1040 TGGTCGTTG ACGACGCAAC	1050 GCCACCCCAAG CGGTGGGTTT	1060 GGTCTTAAG CCAGGAATTC	1070 ACAACCTGTG TGTGGACAC	1080 GCGGTATTCG CGGCATAGC	1090 AACGTATATA TTGCATATAT	1100 GGAACGTTTT CCTTGCAAAA
1110 TCGCAGGGTA AGCGTCCCAT	1120 CTCGAGGACT GAGCTCCTGA	1130 TTCCAAAAAA AAGGTTTTTT	1140 CCCTGCTTTT GGGACGAAAA	1150 TCAGTCGACG AGTCAGCTGC	1160 TTTGGCGCGA AAACCGCGCT	1170 AAAGAAGCAA TTTCTCGGTT	1180 CTCGGTACAA GAGCCATGTT	1190 CACGTATATA GTGCATATTT	1200 AGAACAATGT TCCTGTACAC
1210 CGACGGGACT GCTGCGCTGA	1220 AGGCGGTTTT TCCGGCAAAA	1230 TTGCTTTTCA AACGAAAGGT	1240 CCGCGCGAGC GGCGGCTCG	1250 AGCACTAGGT TCGTGATCCA	1260 GTACCATGAA CATGGTACTT	1270 AAGGTAAATC TTCCATTAGC	1280 TATCGGTACC ATAGCATAGG	1290 GAAGGTACGG CTTCATGCCC	1300 AAAAAGGTTT TTTTTCCAAA
1310 CGACTTTGAT GCTGAACATA	1320 CCCCGACCGA GGGCGTGGCT	1330 ACGGCCTAAC TGCCTGATTT	1340 GCTTGTGTTT CGAACAAACA	1350 GTTGTAAGAA CAACATTCTT	1360 AAGTAAAGC TTCATTTTTC	1370 AGCGACAAAA TCGCTGTTTT	1380 CTCGCCTTCG GAGCGGAAGC	1390 GAAGTTTTTC CTTCAAAACG	1400 ACATGGACGG TGTACCTGCC
1410 ACCAAGGATA TGGTTTCCAT	1420 AAAAAGTTTT TTTTTGAAAA	1430 CTGACTCTTG GACTGAGAAC	1440 CGAGACGTAC CGTCTGCATG	1450 TACGTATTAC ATGCATATAT	1460 GCCTGCGCGC CGGACGGGCG	1470 CGTACGACTT GCATGCTGAA	1480 TGGGTATGTA ACCCATTACT	1490 GGATTTTGAC CCTAAACATG	1500 GAGAACCAAC CTCTTGGTGG
1510 AAGACGGADA TTCCTGCTCT	1520 AGAAGAGAC TCTTCTCTCT	1530 GTGAGAGACC CACTCTCTCT	1540 CCTTCTCTCA GGAAAGAGGT	1550 TAGCGTCGGT ATCGACGACA	1560 ATCTAAAGAA TAGATTCTTT	1570 CTGAAAAAAG GACTTTTTTC	1580 AAACCTCCAT TTTGGAGGTA	1590 TTCCGTGTCC AAGGCACAGC	1600 AAGGTCAAGA TTCCAGTTCT
1610 AGAAGCGAAA TCCTCGCTTT	1620 GCCTTAGGTC CGGAATCCAG	1630 TTTCATAGAC AAAGTATCTG	1640 GGGTAAAGC CCCATTTTTC	1650 CGCGCGCGCG CGCGCGCGCG	1660 GACTCGCGAC CTGAGCGCTG	1670 GCCAGACCCC CGGTCTGGGG	1680 ACGCGAGGGA TGCCTCTCCT	1690 GACACTCAGC CTGTGAGTGC	1700 ACTAACGACC TGATTGCTGG
1710 GCTAATAAAT CCATTATTTA	1720 TAGGATCCGT ATCTAGGCA	1730 TCTTTTGTGT AAGAACAACA	1740 ACTACTAGAA TGATGATCT	1750 CCTCGGTGTC GGAGCCACAG	1760 CTTTCGAATT GAAAGCTTAA	1770 GGCGGGGGTG CCGCCCCCAC	1780 CGGAGGCGCG CGCTCCGCC	1790 TAACCGCCAT ATTGGCGCTA	1800 GCCGACAGTA CGGCTGTCTA
1810 CCTCTTCCCT GGAGAAGGAC	1820 TTTTAGATG AAAAGTCTAC	1830 AGTATGGGGT TCATACCCCA	1840 TCTGCGTGGC AGACGACCG	1850 CAACTCCTCT GTTGAGCAGA	1860 TGAACCCGAT ACTTGGGCTA	1870 GCTCTGAGGG CGAGACTCC	1880 GGGCTCCTTA CCCGAGGAAT	1890 AACTTCCGAA TTGAAGGCTT	1900 AGAAGTTTAG TCTTCAATCT
1910 GTTTTCTGTT CAAAAGCAAC	1920 GTTTACTCGT CAATGAGCA	1930 TTTGGGACCC AAACGCTGGG	1940 GAGCTC CTCGAG						

FIGURE 28.1 (Continued)

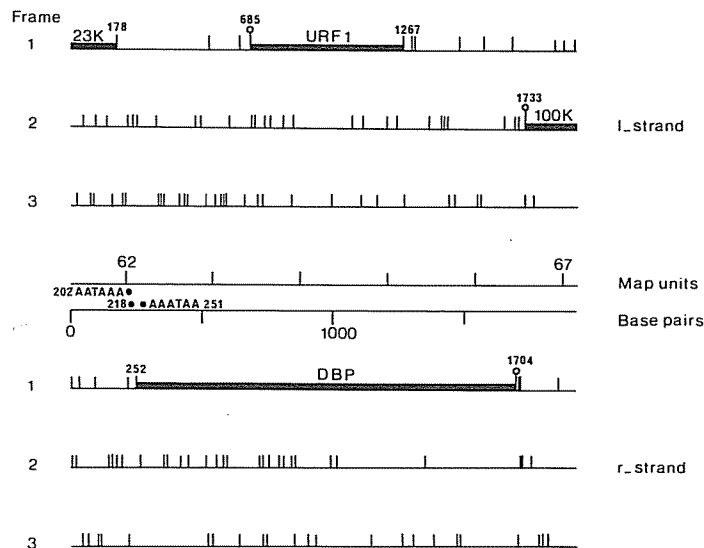


FIGURE 28.2. Structural organization of a region between coordinates 61.5 and 67.0 on the Ad12 genome. This map is derived from the nucleotide sequence in Fig. 28.1.

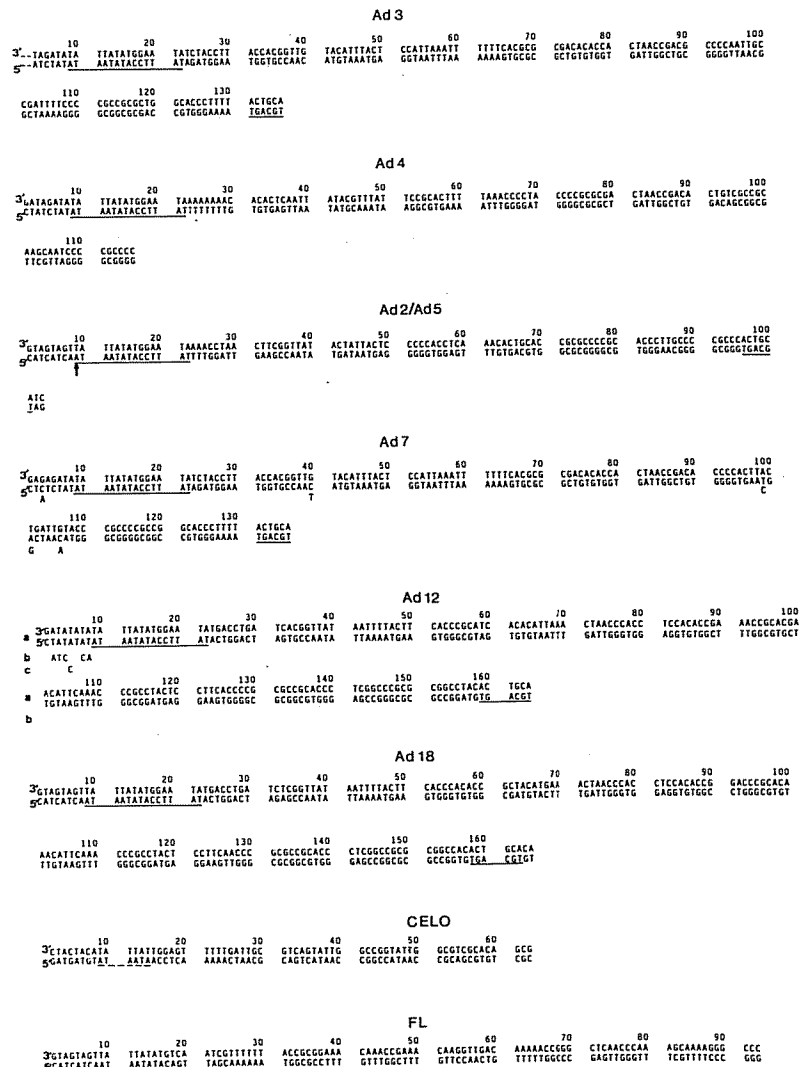


FIGURE 29A, B. Nucleotide sequence of inverted terminal repetitions. The origins of the sequences are as follows: [A] Ad3: Tolun *et al.* (1979). Ad4: Tokunaga *et al.* (1982). Ad2/Ad5: The Ad2 sequence was determined by Shinagawa and Padmanabhan (1979) and the Ad5 sequence by Steenbergh *et al.* (1977). The two sequences are identical. Arrand and Roberts (1979) have analyzed an Ad2 strain that missed base pair 9 (↑). Ad7: These sequences were determined for strain Gomen by Dijkema and Dekker (1979) (a) and for strain Greider by Shinagawa and Padmanabhan (1980) (b). The differences between the sequences are indicated. [B] Ad12: Tolun *et al.* (1979) (a), Sugisaka *et al.* (1980) (a), Shinagawa and Padmanabhan (1980) (b), and Schwarz *et al.* (1982) (c). The differences between the sequences are indicated. Ad18: Garon *et al.* (1982). CELO: Aleström *et al.* (1982a). FL: Temple *et al.* (1981). In the human sequences, the conserved sequences 9–22 are underlined, the homologous regions in CELO and FL DNA are indicated by dashed underlines. The common sequence TGACGT discovered by Shinagawa and Padmanabhan (1980) is underlined.

REFERENCES

- Akusjärvi, G., and Persson, H., 1981a, Gene and mRNA for precursor polypeptide VI from adenovirus type 2, *J. Virol.* **38**:469.
- Akusjärvi, G., and Persson, H., 1981b, Control of RNA splicing and termination in the major late adenovirus transcription unit, *Nature (London)* **292**:420.
- Akusjärvi, G., and Pettersson, U., 1978a, Sequence analysis of adenovirus DNA. I. Nucleotide sequence at the carboxy-terminal end of the gene for adenovirus type 2 hexon, *Virology* **91**:477.
- Akusjärvi, G., and Pettersson, U., 1978b, Nucleotide sequence at the junction between the coding region of the adenovirus 2 hexon messenger RNA and its leader sequence, *Proc. Natl. Acad. Sci. U.S.A.* **75**:5822.
- Akusjärvi, G., and Pettersson, U., 1979a, Sequence analysis of adenovirus DNA: Complete nucleotide sequence of the spliced 5'-non-coding region of adenovirus 2 hexon messenger RNA, *Cell* **16**:841.
- Akusjärvi, G., and Pettersson, U., 1979b, Sequence analysis of adenovirus DNA. IV. The genomic sequences encoding the common tripartite leader of late adenovirus messenger RNA, *J. Mol. Biol.* **13**:143.
- Akusjärvi, G., Mathews, M.B., Anderson, P., Vennström, B., and Pettersson, U., 1980, Structure of genes for virus-associated RNAI and RNAII of adenovirus type 2, *Proc. Natl. Acad. Sci. U.S.A.* **77**:2424.
- Akusjärvi, G., Zabielsky, J., Perricaudet, M., and Pettersson, U., 1981, The sequence of the 3' non-coding region of the hexon mRNA discloses a novel adenovirus gene, *Nucleic Acids Res.* **9**:1.
- Aleström, P., Akusjärvi, G., Perricaudet, M., Mathews, M.B., Klessig, D., and Pettersson, U., 1980, Sequence analysis of adenovirus DNA. VI. The nucleotide sequence of the gene for polypeptide IX from adenovirus type 2 and its colinear messenger RNA, *Cell* **19**:671.
- Aleström, P., Stenlund, A., Li, P., and Pettersson, U., 1982a, A common sequence in the inverted terminal repetitions of human and avian adenoviruses, *Gene* **18**:193.
- Aleström, P., Akusjärvi, G., Pettersson, M., and Pettersson, U., 1982b, DNA sequence analysis of the region encoding the terminal protein and the hypothetical N-gene product of adenovirus type 2, *J. Biol. Chem.* **257**:13492.
- Anderson, C.W., and Lewis, J.B., 1980, Amino-terminal sequence of the adenovirus type 2 proteins: Hexon, fiber, component IX and the early protein 1B-15 K, *Virology* **104**:27.
- Ariga, H., Shimojo, H., Hidaka, S., and Miura, M., 1979, Specific cleavage of the terminal protein from the adenovirus 5 DNA under the conditions of single-strand scission by nuclease S1, *FEBS Lett.* **107**:355.
- Ariga, H., Klein, H., Levine, A.J., and Horwitz, M.S., 1980, A cleavage product of the adenovirus DNA binding protein is active for DNA replication *in vitro*, *Virology* **101**:307.
- Arrand, J.R., and Roberts, R.J., 1979, The nucleotide sequences at the termini of adenovirus 2 DNA, *J. Mol. Biol.* **128**:577.
- Arrand, J.R., Keller, W., and Roberts, R.J., 1975, Extent of terminal repetition in adenovirus 2 DNA, *Cold Spring Harbor Symp. Quant. Biol.* **39**:401.
- Baker, C.C., and Ziff, E.B., 1980, Biogenesis, structures, and sites of encoding of the 5' termini of adenovirus-2 mRNAs, *Cold Spring Harbor Symp. Quant. Biol.* **44**:415.
- Baker, C.C., and Ziff, E.B., 1981, Promoters and heterogeneous 5'-termini of the messenger RNAs of adenovirus-2, *J. Mol. Biol.* **149**:189.
- Baker, C.C., Hérisse, J., Courtois, G., Galibert, F., and Ziff, E., 1979, Messenger RNA for the Ad2 DNA binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5' end, *Cell* **8**:569.
- Bartok, K., Garon, C.F., Berry, K.W., and Rose, J.A., 1974, Specific fragmentation of adenovirus heteroduplex DNA molecules with single-strand specific nuclease of *Neurospora crassa*, *J. Mol. Biol.* **87**:437.

- Bégin, M., and Weber, J., 1975, Genetic analysis of adenovirus type 2. I. Isolation and genetic characterization of temperature-sensitive mutants, *J. Virol.* **115**:1.
- Berget, S.M., Moore, C., and Sharp, P.A., 1977, Spliced segments at the 5' terminus of adenovirus 2 late mRNA, *Proc. Natl. Acad. Sci. U.S.A.* **74**:3171.
- Berget, S.M., Berk, A., Harrison, T., and Sharp, P.A., 1978, Spliced segments at the 5' termini of adenovirus 2 late mRNA: A role for heterogeneous nuclear RNA in mammalian cells, *Cold Spring Harbor Symp. Quant. Biol.* **42**:523.
- Berk, A.J., and Sharp, P.A., 1977a, Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids, *Cell* **12**:721.
- Berk, A.J., and Sharp, P.A., 1977b, UV mapping of the adenovirus 2 early promoters, *Cell* **12**:45.
- Berk, A.J., and Sharp, P.A., 1978, Structure of the adenovirus 2 early mRNAs, *Cell* **14**:695.
- Berk, A.J., Lee, F., Harrison, T., Williams, J., and Sharp, P.A., 1979, Pre-early adenovirus 5 gene product regulates synthesis of early viral messenger RNA, *Cell* **17**:935.
- Bernards, R., Houweling, A., Schrier, P.I., Bos, J.L., and van der Eb, A.J., 1982, Characterization of cells transformed by Ad5/Ad12 hybrid early region 1 plasmids, *Virology* **120**:422.
- Bhatti, A.R., and Weber, J., 1979, Protease of adenovirus type 2, *J. Biol. Chem.* **254**:12265.
- Binger, M.H., Flint, S.J., and Rekosh, D.M., 1982, Expression of the gene encoding the adenovirus DNA terminal protein precursor in productively infected and transformed cells, *J. Virol.* **42**:488.
- Bos, J.L., Polder, L.J., Bernards, R., Schrier, P.I., van den Elsen, P., Van der Eb, A.J., and van Ormondt, H., 1981, The 2.2 kb mRNA of human Ad12 and Ad5 codes for two tumor antigens starting at different AUG triplets, *Cell* **27**:121.
- Brackman, K.H., Green, M., Wold, W.S.M., Cartas, M., Matsuo, T., and Hashimoto, S., 1980, Identification and peptide mapping of human adenovirus type 2-induced early polypeptides isolated by two-dimensional gel electrophoresis, *J. Biol. Chem.* **255**:6772.
- Broker, T.R., Chow, L.T., Dunn, A.R., Gelin, R.E., Hassell, J.A., Klessig, D.F., Lewis, J.B., Roberts, R.J., and Zain, B.S., 1977, Adenovirus-2 messengers—An example of baroque molecular architecture, *Cold Spring Harbor Symp. Quant. Biol.* **42**:531.
- Brown, D.T., Westphal, M., Burlingham, B.T., Winterhoff, U., and Doerfler, W., 1975, Structure and composition of the adenovirus type 2 core, *J. Virol.* **16**:366.
- Burnett, J.P., and Harrington, J.A., 1968, Simian adenovirus SA7 DNA: Chemical, physical and biological studies, *Proc. Natl. Acad. Sci. U.S.A.* **60**:1023.
- Carter, T.H., and Blanton, R.A., 1978, Possible role of the 72,000-dalton DNA-binding protein in regulation of adenovirus type 5 early gene expression, *J. Virol.* **25**:664.
- Carter, T.H., and Ginsberg, H.S., 1976, Viral transcription in KB cells infected by temperature-sensitive early mutants of adenovirus type 5, *J. Virol.* **18**:156.
- Carusi, E.A., 1977, Evidence for blocked 5' termini in human adenovirus DNA, *Virology* **76**:380.
- Celma, M.L., Pan, J., and Weissmann, S., 1977a, Studies of low molecular weight RNA from cells infected with adenovirus 2. I. The sequence at the 3' end of VA-RNA 1, *J. Biol. Chem.* **252**:9032.
- Celma, M.L., Pan, J., and Weissmann, S., 1977b, Studies of low molecular weight RNA from cells infected with adenovirus 2. II. Heterogeneity at the 5' end of VA-RNA 1, *J. Biol. Chem.* **252**:9043.
- Challberg, M.D., and Kelly, T.J., Jr., 1979a, Adenovirus DNA replication *in vitro*, *Proc. Natl. Acad. Sci. U.S.A.* **76**:655.
- Challberg, M.D., and Kelly, T.J., Jr., 1979b, Adenovirus DNA replication *in vitro*: Origin and direction of daughter strand synthesis, *J. Mol. Biol.* **135**:999.
- Challberg, M.D., and Kelly, T.J., Jr., 1981, Processing of the adenovirus terminal protein, *J. Virol.* **38**:272.
- Challberg, M.D., Desiderio, S.V., and Kelly, T.J., Jr., 1980, Adenovirus DNA replication *in vitro*: Characterization of a protein covalently linked to nascent DNA strands, *Proc. Natl. Acad. Sci. U.S.A.* **77**:5105.

- Chinnadurai, G., Chinnadurai, S., and Green, M., 1978, Enhanced infectivity of adenovirus DNA and a DNA-protein complex, *J. Virol.* 26:195.
- Chow, L.T., and Broker, T.R., 1978, The spliced structures of adenovirus-2 fiber message and the other mRNAs, *Cell* 15:497.
- Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J., 1977a, An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA, *Cell* 12:1.
- Chow, L.T., Roberts, J.M., Lewis, J.B., and Broker, T.R., 1977b, A cytoplasmic RNA transcript map from adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids, *Cell* 11:819.
- Chow, L.T., Broker, T.R., and Lewis, J.B., 1979a, Complex splicing patterns of RNAs from the adenovirus 2, *J. Mol. Biol.* 134:265.
- Chow, L., Lewis, J.B., and Broker, T., 1979b, RNA transcription and splicing at early and intermediate times after adenovirus-2 infection, *Cold Spring Harbor Symp. Quant. Biol.* 44:401.
- Coombs, D.H., and Pearson, G.D., 1978, Filter-binding assay for covalent DNA-protein complexes: Adenovirus DNA-terminal protein complex, *Proc. Natl. Acad. Sci. U.S.A.* 75:5291.
- Coombs, D.H., Robinson, A.J., Bodnar, J.W., Jones, C.J., and Pearson, G.D., 1978, Detection of DNA-protein complexes: The adenovirus DNA-terminal protein and HeLa DNA-protein complexes, *Cold Spring Harbor Symp. Quant. Biol.* 43:741.
- Corden, J., Engelking, H.M., and Pearson, G.D., 1976, Chromatin-like organization of the adenovirus chromosome, *Proc. Natl. Acad. Sci. U.S.A.* 73:401.
- Desiderio, S.V., and Kelly, T., 1981, Structure of the linkage between adenovirus DNA and the 55,000 molecular weight terminal protein, *J. Mol. Biol.* 145:319.
- Dijkema, R., and Dekker, B.M.M., 1979, The inverted terminal repetition of the DNA of the weakly oncogenic adenovirus type 7, *Gene* 8:7.
- Dijkema, R., Dekker, B.M.M., and van Ormondt, H., 1980a, The nucleotide sequence of the transforming BglII-H fragment of adenovirus type 7 DNA, *Gene* 9:141.
- Dijkema, R., Dekker, B.M.M., van Ormondt, H., Maat, J., and Boyer, H., 1980b, Gene organization of the transforming region of weakly oncogenic adenovirus type 7: The E1a region, *Gene* 12:287.
- Dijkema, R., Maat, J., Dekker, B.M.M., van Ormondt, H., and Boyer, W., 1981, The gene for polypeptide IX of adenovirus type 7, *Gene* 13:375.
- Dijkema, R., Dekker, B.M.M., and van Ormondt, H., 1982, Gene organization of the transforming region of adenovirus type 7, *Gene* 18:143.
- Dimitrov, D.H., Dubitchew, A.G., Naroditsky, B.S., Dreizin, R.S., and Tikchonenko, T.I., 1979, Physicochemical properties and restriction maps of simian adenovirus 38, *J. Gen. Virol.* 44:69.
- Doerfler, W., and Kleinschmidt, A., 1970, Denaturation pattern of the DNA of adenovirus type 2 as determined by electron microscopy, *J. Mol. Biol.* 50:579.
- Doerfler, W., Hellmann, W., and Kleinschmidt, A.K., 1972, The DNA of adenovirus type 12 and its denaturation pattern, *Virology* 47:507.
- Dunn, A.R., Mathews, M.B., Chow, L.T., Sambrook, J., and Keller, W., 1978, A supplementary adenoviral leader sequence and its role in messenger translation, *Cell* 15:511.
- Ellens, D.J., Sussenbach, J.S., and Jansz, H.S., 1974, Studies on the mechanism of replication of adenovirus DNA. III. Electron microscopy of replicating DNA, *Virology* 61:427.
- Engler, J., 1981, The nucleotide sequence of the polypeptide IX gene of human adenovirus type 3, *Gene* 13:387.
- Engler, J.A., and van Bree, M., 1982, The nucleotide sequence of the gene encoding protein IVa2 in human adenovirus type 7, *Gene* 19:71.
- Engler, J., Broker, T.R., and Chow, L.T., 1981, Sequences of human adenovirus Ad3 and Ad7 DNAs encoding the promoter and first leader segment of late RNAs, *Gene* 13:133.
- Engler, J.A., Hoppe, M.S., and van Bree, M.P., 1983, The nucleotide sequence of the genes encoded in the early region 2B in human adenovirus type 7, *Gene* 21:147.
- Enomoto, T., Lichy, J.H., Ikeda, J.E., and Hurwitz, J., 1981, Adenovirus DNA replication in

- vitro*: Purification of the terminal protein in a functional form, *Proc. Natl. Acad. Sci. U.S.A.* 78:6779.
- Esche, H., Mathews, M.B., and Lewis, J.B., 1980, Proteins and messenger RNAs of the transforming region of wild-type and mutant adenoviruses, *J. Mol. Biol.* 142:399.
- Evans, R.M., Fraser, N., Ziff, E., Weber, J., Wilson, M., and Darnell, J.E., 1977, The initiation sites for RNA transcription in Ad2 DNA, *Cell* 12:733.
- Everitt, E., Sundquist, B., Petersson, O., and Philipson, L., 1973, Structural proteins of adenoviruses. X. Isolation and topography of low molecular weight antigens from the virion of adenovirus type 2, *Virology* 62:130.
- Forsblom, S., Rigler, R., Ehrenberg, M., Pettersson, U., and Philipson, L., 1976, Kinetic studies on the cleavage of adenovirus DNA by restriction endonuclease EcoRI, *Nucleic Acids Res.* 3:3255.
- Fraser, N., and Ziff, E., 1978, RNA structures near poly (A) of adenovirus-2 late messenger RNA's, *J. Mol. Biol.* 124:27.
- Fraser, N.W., Nevins, J.R., Ziff, E., and Darnell, J.E., 1979, The major late adenovirus type 2 transcription unit: Termination is downstream from the last poly (A) site, *J. Mol. Biol.* 129:643.
- Fraser, N.W., Baker, C.C., Moore, M.A., and Ziff, E.B., 1982, Poly(A) sites of adenovirus serotype 2 transcription units, *J. Mol. Biol.* 155:207.
- Freeman, A.E., Black, P.E., Vanderpool, E.A., Henry, P.H., Austin, J.B., and Huebner, R.J., 1967, Transformation of primary rat embryo cells by adenovirus type 2, *Proc. Natl. Acad. Sci. U.S.A.* 58:1205.
- Frost, E., and Williams, J., 1978, Mapping of temperature-sensitive and host-range mutants of adenovirus type 5 by marker rescue, *Virology* 91:39.
- Fujinaga, K., Sawada, Y., Uemizu, Y., Yamashita, T., Shimojo, H., Shiroki, K., Sugisaki, H., Sugimoto, K., and Takanami, M., 1979, Nucleotide sequences, integration, and transcription of the adenovirus-12 transforming genes, *Cold Spring Harbor Symp. Quant. Biol.* 44:519.
- Galibert, F., Hérissey, J., and Courtois, G., 1979, Nucleotide sequence of the EcoRI-F fragment of adenovirus 2 genome, *Gene* 6:1.
- Gallimore, P.H., 1974, Interactions of adenovirus type 2 with rat embryo cells: Permissiveness, transformation and *in vitro* characteristics of adenovirus transformed rat embryo cells, *J. Gen. Virol.* 25:263.
- Galos, R.S., Williams, J., Binger, M.H., and Flint, S.J., 1979, Location of additional early gene sequences in the adenoviral chromosome, *Cell* 17:945.
- Gambke, C., and Deppert, W., 1981, Late non-structural 100,000 and 33,000 dalton proteins of adenovirus type 2. II. Immunological and protein chemical analysis, *J. Virol.* 40:594.
- Garon, C.F., Berry, K.W., and Rose, J.A., 1972, A unique form of terminal redundancy in adenovirus DNA molecules, *Proc. Natl. Acad. Sci. U.S.A.* 69:2391.
- Garon, C.F., Berry, K.W., Hierholzer, J.C., and Rose, J.A., 1973, Mapping base sequence heterologies between genomes from different adenovirus serotypes, *Virology* 54:414.
- Garon, C.F., Berry, K.W., and Rose, J.A., 1975, Arrangement of sequences in the inverted terminal repetition of adenovirus 18 DNA, *Proc. Natl. Acad. Sci. U.S.A.* 72:3039.
- Garon, F.C., Parr, R.P., Padmanabhan, R., Roninson, I., Garrison, J.W., and Rose, J.A., 1982, Structural characterization of the adenovirus 18 inverted terminal repetition, *Virology* 121:230.
- Gingeras, T.R., Sciaky, D., Gelinas, R.E., Bing-Dong, J., Yen, C.E., Kelly, M.M., Bullock, P.A., Parsons, B.L., O'Neill, K.E., and Roberts, R.J., 1982, Nucleotide sequences from the adenovirus-2 genome, *J. Biol. Chem.* 257:13475.
- Ginsberg, H.S., Ensinger, M.J., Kauffman, R.S., Mayer, A.J., and Lundholm, U., 1974, Cell transformation: A study of regulation with types 5 and 12 adenovirus temperature-sensitive mutants, *Cold Spring Harbor Symp. Quant. Biol.* 44:419.
- Girardi, A.J., Hilleman, M.R., and Zwick, R.E., 1964, Tests in hamsters for oncogenic quality of ordinary viruses including adenovirus type 7, *Proc. Soc. Exp. Biol. Med.* 115:1141.

- Goldbach, R.W., Evers, R.F., and Borst, P., 1978, Electrophoretic strand separation of long DNAs with poly-UG in agarose gels, *Nucleic Acids Res.* 5:2743.
- Goodhearst, C.R., 1971, DNA density of oncogenic and non-oncogenic simian adenoviruses, *Virology* 44:645.
- Graham, F.L., and van der Eb, A.J., 1973, A new technique for the assay of the infectivity of human adenovirus 5 DNA, *Virology* 52:456.
- Graham, F.L., van der Eb, A.J., and Heijneker, H.L., 1974a, Size and location of the transforming regions of human adenovirus type 5 DNA, *Nature (London)* 251:687.
- Graham, F.L., Abrahams, P.J., Mulder, C., Heijneker, H.L., Warnaar, S.O., de Vries, F.A.J., Fiers, W., and van der Eb, A.J., 1974b, Studies on *in vitro* transformation by DNA and DNA fragments of human adenoviruses and simian virus 40, *Cold Spring Harbor Symp. Quant. Biol.* 39:637.
- Green, M., 1970, Oncogenic viruses, *Annu. Rev. Biochem.* 39:701.
- Green, M., and Piña, M., 1963, Biochemical studies on adenovirus multiplication. IV. Isolation, purification and chemical analysis of adenovirus, *Virology* 20:199.
- Green, M., and Piña, M., 1964, Biochemical studies on adenovirus multiplication. VI. Properties of highly purified tumorigenic human adenoviruses and their DNA's, *Proc. Natl. Acad. Sci. U.S.A.* 51:1251.
- Green, M., Piña, M., Kimes, R., Wensink, P.C., Machattie, L.A., and Thomas, C.A., 1967, Adenovirus DNA. I. Molecular weight and conformation, *Proc. Natl. Acad. Sci. U.S.A.* 57:1302.
- Green, M., Parsons, J.T., Piña, M., Fujinaga, K., Caffier, H., and Landgraf-Leurs, I., 1970, Transcription of adenovirus genes in productively infected and transformed cells, *Cold Spring Harbor Symp. Quant. Biol.* 30:803.
- Green, M., Wold, W.S.M., Brackmann, K.H., and Cartas, M.A., 1979a, Identification of families of overlapping polypeptides coded by early "transforming" gene region 1 of human adenovirus type 2, *Virology* 97:275.
- Green, M., Mackey, J., Wold, W.S.M., and Rigden, P., 1979b, Thirty-one human adenovirus serotypes (Ad1-Ad31) form five groups (A-E) based on upon DNA genome homologies, *Virology* 93:481.
- Green, M., Brackmann, K., Wold, W.S.M., Cartas, M., Thornton, H., and Elder, J.H., 1979c, Conserved primary sequences of the DNA terminal proteins of five different human adenovirus groups, *Proc. Natl. Acad. Sci. U.S.A.* 76:4380.
- Green, M., Wold, W.S.M., Brackmann, K., and Cartas, M.A., 1979d, Studies on early proteins and transformation proteins in human adenoviruses, *Cold Spring Harbor Symp. Quant. Biol.* 44:457.
- Grodzicker, T., Williams, J., Sharp, P.A., and Sambrook, J., 1975, Physical mapping of temperature-sensitive mutations of adenoviruses, *Cold Spring Harbor Symp. Quant. Biol.* 39:439.
- Grodzicker, T., Anderson, C.W., Sambrook, J., and Mathews, M.B., 1977, The physical location of structural genes in adenovirus DNA, *Virology* 80:111.
- Harrison, T., Graham, F.L., and Williams, J., 1977, Host range mutants of adenovirus type 5 defective for growth in HeLa cells, *Virology* 77:319.
- Harter, M.L., and Lewis, J.B., 1978, Adenovirus type 2 early proteins synthesized *in vitro* and *in vivo*: Identification in infected cells of the 38,000- to 50,000-molecular-weight protein encoded by the left end of adenovirus type 2 genome, *J. Virol.* 26:736.
- Harter, M.L., Shanmugan, G., Wold, W.S.M., and Green, M., 1976, Detection of adenovirus type 2 induced early polypeptides using cycloheximide pretreatment to enhance viral protein synthesis, *J. Virol.* 19:1976.
- Hashimoto, S., Pursley, M.H., and Green, M., 1981, Nucleotide sequences and mapping of novel heterogeneous 5'-termini of adenovirus type 2 early region 4 mRNA, *Nucleic Acids Res.* 9:1675.
- Hérissé, J., and Galibert, F., 1981, Nucleotide sequence of the EcoRI E fragment of adenovirus-2 genome, *Nucleic Acids Res.* 9:1229.
- Hérissé, J., Courtois, G., and Galibert, F., 1980, Nucleotide sequence of the EcoRI D fragment of adenovirus 2 genome, *Nucleic Acids Res.* 8:2173.

- Hérissé, J., Rigolet, M., Dupont de Dinechin, S., and Galibert, F., 1981, Nucleotide sequence of adenovirus 2 fragment encoding for the carboxylic region of the fiber protein and the entire E4 region, *Nucleic Acids Res.* **9**:4023.
- Hierholzer, J.C., 1973, Further subgrouping of the human adenoviruses by differential hemagglutination, *J. Infect. Dis.* **128**:541.
- Horwitz, M.S., 1974, Location of the origin of DNA replication in adenovirus type 2, *J. Virol.* **13**:1046.
- Horwitz, M.S., 1978, Temperature-sensitive replication of H5ts125 adenovirus DNA in the presence of cycloheximide, *J. Virol.* **11**:544.
- Houweling, A., van den Elsen, P.J., and van der Eb, A.J., 1980, Partial transformation of primary rat cells by the leftmost 4.5% fragment of adenovirus 5 DNA, *Virology* **105**:537.
- Huebner, R.J., Rowe, W.P., and Lane, W.T., 1962, Oncogenic effects in hamsters of human adenovirus types 12 and 18, *Proc. Natl. Acad. Sci. U.S.A.* **48**:2051.
- Huebner, R.J., Casey, M.J., Chanock, R.M., and Schell, K., 1965, Tumors induced in hamsters by a strain of adenovirus type 3: Sharing of tumor antigens and neoantigens with those produced by adenovirus type 7 tumors, *Proc. Soc. Exp. Biol. Med.* **54**:381.
- Jones, N., and Shenk, T., 1979a, An adenovirus type 5 early gene function regulates expression of other early viral genes, *Proc. Natl. Acad. Sci. U.S.A.* **76**:3665.
- Jones, N., and Shenk, T., 1979b, Isolation of adenovirus type 5 host range deletion mutants defective for transformation of rat embryo cells, *Cell* **17**:683.
- Jörnvall, H., Akusjärvi, G., Aleström, P., von Bahr-Lindstrom, H., Pettersson, U., Appella, E., Fowler, A.V., and Philipson, L., 1981a, The adenovirus hexon protein: The primary structure of the polypeptide and its correlation with the hexon gene, *J. Biol. Chem.* **256**:6181.
- Jörnvall, H., Aleström, P., Akusjärvi, G., von Bahr-Lindstrom, H., Philipson, L., and Pettersson, U., 1981b, Order of the CNBr fragments in the adenovirus hexon protein, *J. Biol. Chem.* **256**:6204.
- Keegstra, W.S., van Wielink, P.S., and Sussenbach, J.S., 1977, The visualization of a circular DNA-protein complex from adenovirions, *Virology* **76**:444.
- Kilpatrick, B.A., Gelinas, R.E., Broker, T.R., and Chow, L.T., 1979, Comparison of late mRNA splicing among class B and class C adenoviruses, *J. Virol.* **30**:899.
- Kimes, R., and Green, M., 1970, Adenovirus DNA. II. Separation of molecular halves of adenovirus type 2, *J. Mol. Biol.* **50**:203.
- Kimura, T., Sawada, Y., Shinagawa, M., Shimizu, Y., Shiroki, K., Shimojo, H., Sugisaki, H., Takanami, M., Uemizu, Y., and Fujinaga, K., 1981, Nucleotide sequence of the transforming early region E1b of adenovirus 12 DNA: Structure and gene organization and comparison with those of adenovirus type 5 DNA, *Nucleic Acids Res.* **9**:6571.
- Kitchingman, G.R., and Westphal, H., 1980, The structure of adenovirus 2 early nuclear and cytoplasmic RNAs, *J. Mol. Biol.* **137**:23.
- Kitchingman, G.R., Lai, S.-P., and Westphal, H., 1977, Loop structures in hybrids of early RNA and the separated strands of adenovirus DNA, *Proc. Natl. Acad. Sci. U.S.A.* **74**:4329.
- Klessig, D.F., and Grodzicker, T., 1979, Mutations that allow human Ad2 and Ad5 to express late genes in monkey cells map in the viral gene encoding the 72K DNA binding protein, *Cell* **17**:957.
- Kruijer, W., van Schaik, F.M.A., and Sussenbach, J.S., 1980, Nucleotide sequence analysis of a region of adenovirus 5 DNA encoding a hitherto unidentified gene, *Nucleic Acids Res.* **8**:6033.
- Kruijer, W., van Schaik, F.A.M., and Sussenbach, J.S., 1981, Structure and organization of the gene coding for the DNA binding protein of adenovirus type 5, *Nucleic Acids Res.* **9**:4439.
- Kruijer, W., van Schaik, F.A.M., and Sussenbach, J.S., 1982, Nucleotide sequence of the gene encoding adenovirus type 2 DNA binding protein, *Nucleic Acids Res.* **10**:4493.
- Kruijer, W., van Schaik, F.A.M., Speyer, J.G., and Sussenbach, J.S., 1983, Structure and function of adenovirus DNA binding protein: comparison of the amino acid sequences of

- Ad5 and Ad12 proteins derived from the nucleotide sequence of the corresponding genes, *Virology* 128:140.
- Kubinski, H., and Rose, J.A., 1967, Regions containing repeating base-pairs in DNA from some oncogenic and non-oncogenic animal viruses, *Proc. Natl. Acad. Sci. U.S.A.* 57:1720.
- Landgraf-Leurs, M., and Green, M., 1971, Adenovirus DNA. III. Separation of the complementary strands of adenovirus types 2, 7 and 12 DNA molecules, *J. Mol. Biol.* 60:185.
- Larsen, S.H., Margolskee, R.F., and Nathans, D., 1979, Alignment of the restriction map of mouse adenovirus FL and human adenovirus type 2, *Virology* 97:406.
- Larson, V.M., Girardi, A.J., Hilleman, M.R., and Zwickey, R.E., 1965, Studies on oncogenicity of adenovirus type 7 viruses in hamsters, *Proc. Soc. Exp. Biol. Med.* 118:15.
- Laver, W.G., 1970, Isolation of an arginine-rich protein from particles of adenovirus type 2, *Virology* 41:488.
- Laver, W.G., Suriano, J.R., and Green, M., 1967, Adenovirus proteins. II. N-terminal amino acid analysis, *J. Virol.* 1:723.
- Laver, W.G., Pereira, H.G., Russell, W.C., and Valentine, R.C., 1968, Isolation of an internal component of adenovirus type 5, *J. Mol. Biol.* 37:379.
- Laver, W.G., Younghusband, H.B., and Wrigley, N.G., 1971, Purification and properties of chick embryo lethal orphan virus (CELO), *Virology* 45:598.
- Lewis, J.B., and Mathews, M.B., 1980, Control of adenovirus early gene expression: A class of "immediate early" products, *Cell* 21:303.
- Lewis, J.B., Atkins, J.F., Anderson, C.W., Baum, P.R., and Gesteland, R.F., 1975, Mapping of late adenovirus genes by cell-free translation of RNA selected by hybridization to specific DNA fragments, *Proc. Natl. Acad. Sci. U.S.A.* 72:1344.
- Lewis, J.B., Atkins, J.F., Baum, P.R., Solem, R., Gesteland, R.B., and Anderson, C.W., 1976, Location and identification of the genes for adenovirus type 2 early polypeptides, *Cell* 17:141.
- Lewis, J.B., Anderson, C.W., and Atkins, J.F., 1977, Further mapping of late adenovirus genes by cell-free translation of RNA selected by hybridization to specific DNA fragments, *Cell* 12:37.
- Lewis, J.B., Esche, H., Smart, J.E., Stillman, B.W., Harter, M.L., and Mathews, M.B., 1979, Organization and expression of the left third of the genome of adenovirus, *Cold Spring Harbor Symp. Quant. Biol.* 44:493.
- Lichy, J.H., Field, J., Horwitz, M.S., and Hurwitz, J., 1982, Separation of the adenovirus terminal protein precursor from its associated DNA polymerase: Role of both proteins in the initiation of adenovirus DNA replication, *Proc. Natl. Acad. Sci. U.S.A.* 79:5225.
- Lupker, J.H., Davis, A., Jochemsen, H., and van der Eb, A.J., 1980, *In vitro* synthesis of adenovirus type 5 T antigens. I. Translation of early region 1-specific RNA from lytically infected cells, *J. Virol.* 37:524.
- Maat, J., and van Ormondt, H., 1979, The nucleotide sequence of the transforming HindIII-G fragment of adenovirus type 5 DNA, *Gene* 6:75.
- Maat, J., van Beveren, C.P., and van Ormondt, H., 1980, The nucleotide sequence of adenovirus type 5 early region E-1: The region between map positions 8.0 (HindIII site) and 11.8 (SmaI site), *Gene* 10:27.
- Maizel, J.V., White, D.O., and Scarff, M.D., 1968, The polypeptides of adenovirus. II. Soluble proteins, cores, top components and structure of the virion, *Virology* 36:126.
- Mathews, M.B., 1980, Binding of adenovirus VA RNA to mRNA: A possible role in splicing?, *Nature (London)* 285:575.
- Mathews, M.B., and Pettersson, U., 1978, The low molecular weight RNA's of adenovirus 2 infected cells, *J. Mol. Biol.* 119:293.
- Mathis, D.J., Elkaim, R., Keding, C., Sassone-Corsi, P., and Chambon, P., 1981, Specific *in vitro* initiation of transcription on the adenovirus type 2 early and late E1I transcription units, *Proc. Natl. Acad. Sci. U.S.A.* 78:7383.
- Mautner, V., Williams, J., Sambrook, J., Sharp, P.A., and Grodzicker, T., 1975, The location of the genes coding for hexon and fibre proteins in adenovirus DNA, *Cell* 5:93.

- Mayer, A.J., and Ginsberg, H.S., 1977, Persistence of type 5 adenovirus DNA in cells transformed by the temperature-sensitive mutant H5ts125, *Proc. Natl. Acad. Sci. U.S.A.* **74**:785.
- McAllister, R.M., Nicolson, M.O., Lewis, A.M., McPherson, I., and Huebner, R.J., 1969, Transformation of rat embryo cells by adenovirus type 1, *J. Gen. Virol.* **4**:29.
- McBride, W.D., and Weiner, A., 1964, *In vitro* transformation of hamster kidney cells by human adenovirus type 12, *Proc. Soc. Exp. Biol. Med.* **115**:870.
- McGrogan, M., and Raskas, H.J., 1978, Two regions of the adenovirus 2 genome specify families of late polysomal RNA's containing common sequences, *Proc. Natl. Acad. Sci. U.S.A.* **75**:625.
- Miller, S.J., Ricciardi, R.P., Roberts, B.E., Paterson, B.M., and Mathews, M.B., 1980, Arrangement of messenger RNAs and protein coding sequences in the major late transcription unit of adenovirus 2, *J. Mol. Biol.* **142**:455.
- Mirza, M.A., and Weber, J., 1982, Structure of adenovirus chromatin, *Biochim. Biophys. Acta* **696**:76.
- Montell, C., Fisher, E.F., Caruthers, M.H., and Berk, A.J., 1982, Resolving the functions of overlapping viral genes by site-specific mutagenesis at a mRNA splice site, *Nature (London)* **295**:380.
- Murray, V., and Holliday, R., 1979, Mechanism of RNA splicing for gene transcripts, *FEBS Lett.* **106**:5.
- Naroditsky, B.S., Kolinina, T.I., Goldberg, E.Z., Borovik, A.S., Karamov, E.V., and Tikchenko, T.I., 1980, Analysis of DNA from human adenovirus type 6 with restriction endonucleases HindIII, BglII, and BamHI, *Biochim. Biophys. Acta* **606**:214.
- Nevins, J.R., 1981, Mechanism of activation of early viral transcription by the adenovirus E1A gene products, *Cell* **26**:213.
- Nevins, J.R., and Darnell, J.E., 1978, Groups of adenovirus type 2 mRNA's derived from a large primary transcript: Probably nuclear origin and possible common 3'-ends, *J. Virol.* **25**:811.
- Nevins, J.R., and Jensen-Winkler, J.J., 1980, Regulation of early adenovirus transcription: A protein product of early region 2 specifically represses region 4 transcription, *Proc. Natl. Acad. Sci. U.S.A.* **77**:1893.
- Nevins, J.R., Blanchard, J.M., and Darnell, J.E., 1980, Transcription units of adenovirus type 2: Termination of transcription beyond the poly(A) addition site in early regions 2 and 4, *J. Mol. Biol.* **144**:377.
- Nicolson, M.O., and McAllister, R.M., 1972, Infectivity of human adenovirus 1 DNA, *Virology* **48**:14.
- Ohe, K., 1972, Virus-coded origin of a low molecular weight RNA from KB cells infected with adenovirus 2, *Virology* **47**:726.
- Ohe, K., and Weissmann, S., 1970, Nucleotide sequence of an RNA from cells infected with adenovirus 2, *Science* **167**:879.
- Ohe, K., and Weissman, S., 1971, The nucleotide sequence of a low molecular weight RNA from cells infected with adenovirus 2, *J. Biol. Chem.* **246**:6999.
- Padmanabhan, R., and Green, M., 1976, Evidence for palindromic sequences near the termini of adenovirus 2 DNA, *Biochem. Biophys. Res. Commun.* **69**:860.
- Padmanabhan, R., and Padmanabhan, R.V., 1977, Specific interaction of a protein(s) near the termini of adenovirus 2 DNA, *Biochem. Biophys. Res. Commun.* **80**:955.
- Pan, J., Celma, M.L., and Weissman, S.M., 1977, Studies of low molecular weight RNA from cells infected with adenovirus 2. III. The sequence of the promoter for VA-RNA I. *J. Biol. Chem.* **252**:9047.
- Patch, C.T., Lewis, A.M., and Levine, A.S., 1972, Evidence for a transcriptional control region of SV40 in the adenovirus 2-SV40 hybrid, Ad2+ND1, *Proc. Natl. Acad. Sci. U.S.A.* **69**:3375.
- Pereira, M.S., Pereira, H.G., and Clarke, S.K., 1965, Human adenovirus type 31: A new serotype with oncogenic properties, *Lancet* **1**:21.
- Perricaudet, M., Akusjärvi, G., Virtanen, A., and Pettersson, U., 1979, Structure of two

- spliced mRNA's from the transforming region of human subgroup C adenoviruses, *Nature (London)* **281**:694.
- Perricaudet, M., le Moullec, J.M., and Pettersson, U., 1980, The predicted structure of two adenovirus T-antigens, *Proc. Natl. Acad. Sci. U.S.A.* **77**:3778.
- Persson, H., and Philipson, L., 1982, Regulation of adenovirus gene expression, *Curr. Top. Microbiol. Immunol.* **97**:157.
- Persson, H., Pettersson, U., and Mathews, M.B., 1978, Synthesis of a structural polypeptide in the absence of viral DNA replication, *Virology* **90**:67.
- Persson, H., Mathisen, B., Philipson, L., and Pettersson, U., 1979a, A maturation protein in adenovirus morphogenesis, *Virology* **93**:198.
- Persson, H., Signas, C., and Philipson, L., 1979b, Purification and characterization of an early glycoprotein from adenovirus type 2 infected cells, *J. Virol.* **29**:938.
- Persson, H., Jansson, M., and Philipson, L., 1980a, Synthesis and genomic site for an adenovirus type 2 early protein, *J. Mol. Biol.* **136**:375.
- Persson, H., Jörnval, H., and Zabielski, J., 1980b, Multiple mRNA species for the precursor to an adenovirus-encoded glycoprotein: Identification and structure of the signal sequence, *Proc. Natl. Acad. Sci. U.S.A.* **77**:6349.
- Pettersson, U., and Mathews, M.B., 1977, The gene and the messenger RNA for adenovirus polypeptide IX, *Cell* **12**:741.
- Piña, M., and Green, M., 1965, Biochemical studies on adenovirus multiplication. IX. Chemical and base composition analysis of 28 human adenoviruses, *Proc. Natl. Acad. Sci. U.S.A.* **54**:547.
- Ponomareva, T.I., Grodnitskaya, N.A., Goldberg, E.E., Chaplygina, N.M., Naroditsky, B.S., and Tikchonenko, T.I., 1979, Biological activity of intact and cleaved DNA of the simian adenovirus 7, *Nucleic Acids Res.* **5**:3119.
- Prage, L., and Pettersson, U., 1971, Structural proteins of adenoviruses. VII. Purification and properties of an arginine-rich core protein from adenovirus type 2 and type 3, *Virology* **45**:364.
- Prage, L., Pettersson, U., and Philipson, L., 1968, Internal base proteins in adenovirus, *Virology* **36**:508.
- Prage, L., Pettersson, U., Hoglund, S., Lonberg-Holm, K., and Philipson, L., 1970, Structural proteins of adenoviruses. IV. Sequential degradation of the adenovirus type 2 virion, *Virology* **42**:341.
- Price, R., and Penman, S., 1972, A distinct RNA polymerase activity, synthesizing 5.5S, 5S and 4S RNA in nuclei from adenovirus 2 infected HeLa cells, *J. Mol. Biol.* **70**:435.
- Rekosh, D., 1981, Analysis of the DNA-terminal protein from different serotypes of human adenovirus, *J. Virol.* **40**:329.
- Rekosh, D.M.K., Russell, W.C., Bellett, A.J.D., and Robinson, A.J., 1977, Identification of a protein linked to the ends of adenovirus DNA, *Cell* **11**:283.
- Rijnders, A.W.M., van Maarschalkerweerd, M.W., Visser, L., Reemst, A.M.C.B., Sussenbach, J.S., and Rozijn, T.H., 1983, Expression of integrated viral DNA sequences outside the transforming region in eight adenovirus-transformed cell lines, *Biochim. Biophys. Acta* **739**:48.
- Roberts, R.J., 1981, Restriction and modification enzymes and their recognition sequences, *Nucleic Acids Res.* **9**:r75.
- Roberts, R.J., Arrand, J.R., and Keller, W., 1974, The length of the terminal repetition in adenovirus-2 DNA, *Proc. Natl. Acad. Sci. U.S.A.* **71**:3829.
- Robinson, A.J., and Bellett, A.J.D., 1975a, A circular DNA-protein complex from adenovirus and its possible role in DNA replication, *Cold Spring Harbor Symp. Quant. Biol.* **39**:523.
- Robinson, A.J., and Bellett, A.J.D., 1975b, Complementary strands of CELO virus DNA, *J. Virol.* **15**:458.
- Robinson, A.J., Younghusband, H.B., and Bellett, A.J.D., 1973, A circular DNA-protein complex from adenoviruses, *Virology* **56**:54.
- Roninson, I., and Padmanabhan, R., 1980, Studies on the nature of the linkage between the terminal protein and the adenovirus DNA, *Biochem. Biophys. Res. Commun.* **94**:398.

- Rósen, L., 1960, A hemagglutination-inhibition technique for typing adenoviruses, *Am. J. Hyg.* 71:120.
- Ross, S., Flint, S.J., and Levine, A.J., 1980, Identification of the adenovirus early proteins and their genomic positions, *Virology* 100:419.
- Russell, W.C., McIntosh, K., and Skehel, J.J., 1971, The preparation and properties of adenovirus cores, *J. Gen. Virol.* 11:35.
- Sambrook, J., Williams, J.F., Sharp, P.A., and Grodzicker, T., 1975, Physical mapping of temperature-sensitive mutations of adenoviruses. *J. Mol. Biol.* 97:369.
- Schwarz, E., Reinke, C., Yamamoto, N., and Zur Hausen, H., 1982, Terminal rearrangements in the genome of adenovirus type 12 mutants adapted to growth in two human tumor cell lines, *Virology* 116:284.
- Seghal, P., Fraser, N., and Darnell, J., 1979, Early Ad2 transcription units: Only promoter-proximal RNA continues to be made in the presence of DRB, *Virology* 94:185.
- Sekikawa, K., Shiroki, K., Shimojo, H., Ojima, S., and Fujinaga, K., 1978, Transformation of a rat cell line by an adenovirus 7 DNA fragment *Virology* 88:1.
- Sharp, P.A., Gallimore, P.H., and Flint, S.J., 1975, Mapping of adenovirus 2 RNA sequences in lytically infected cells and transformed cell lines, *Cold Spring Harbor Symp. Quant. Biol.* 39:457.
- Sharp, P.A., Moore, C., and Haverty, J., 1976, The infectivity of adenovirus 5-DNA protein complex, *Virology* 75:442.
- Shaw, A.R., and Ziff, E.B., 1980, Transcripts from the adenovirus-2 late promoter yields a single family of coterminal mRNA's during early infection and five families at late times, *Cell* 22:905.
- Shinagawa, M., and Padmanabhan, R., 1979, Nucleotide sequences at the inverted terminal repetition of Ad2 DNA, *Biochem. Biophys. Res. Commun.* 87:671.
- Shinagawa, M., and Padmanabhan, R., 1980, Comparative sequence analysis of the inverted terminal repetitions from different adenoviruses, *Proc. Natl. Acad. Sci. U.S.A.* 77:3831.
- Shinagawa, M., Padmanabhan, R.V., and Padmanabhan, R., 1980, The nucleotide sequence of the right-hand terminal Smal-K fragment of adenovirus type 2 DNA, *Gene* 9:99.
- Smart, J.E., and Stillman, B.W., 1982, Adenovirus terminal protein precursor: Partial amino acid sequence and the site of covalent linkage to virus DNA, *J. Biol. Chem.* 257:13499.
- Söderland, H., Pettersson, U., Vennström, B., Philipson, L., and Mathews, M.B., 1976, A new species of virus-coded low molecular weight RNA from cells infected with adenovirus type 2, *Cell* 7:585.
- Spector, D.J., McGrogan, M., and Raskas, H.J., 1978, Regulation of the appearance of cytoplasmic RNA's from region 1 of the adenovirus genome, *J. Mol. Biol.* 126:395.
- Spector, D.J., Crossland, L.D., Halbert, D.N., and Raskas, H.J., 1980a, A 28 K polypeptide is the translation product of 9 S RNA encoded by region 1A of adenovirus 2, *Virology* 102:218.
- Spector, D.J., Halbert, D.N., Crossland, L.D., and Raskas, H.J., 1980b, Expression of genes from the transforming region of adenovirus, *Cold Spring Harbor Symp. Quant. Biol.* 44:437.
- Steenbergh, P.H., and Sussenbach, J.S., 1979, The nucleotide sequence of the right-hand terminus of adenovirus type 5 DNA: Implications for the mechanism of DNA replication, *Gene* 6:307.
- Steenbergh, P.H., Maat, J., van Ormondt, H., and Sussenbach, J.S., 1977, The nucleotide sequence at the termini of adenovirus type 5 DNA, *Nucleic Acids Res.* 4:4371.
- Stenlund, A., Perricaudet, M., Tiollais, P., and Pettersson, U., 1980, Construction of restriction enzyme fragment libraries containing DNA from human adenovirus types 2 and 5, *Gene* 10:47.
- Stillman, B.W., Lewis, J.B., Chow, L.T., Mathews, M.B., and Smart, J.E., 1981, Identification of the gene and mRNA for the adenovirus terminal protein precursor, *Cell* 23:497.
- Sugisaka, H., Sugimoto, K., Takanami, M., Shiroki, K., Saito, I., Shimojo, H., Sawada, Y., Uemizu, Y., Uesigi, S.-I., and Fujinaga, K., 1980, Structure and gene organization in the transforming HindIII-G fragment of Ad12, *Cell* 20:777.
- Sussenbach, J.S., Ellens, D.J., and Jansz, H.S., 1973, Studies on the mechanism of adenovirus

- DNA replication. II. The nature of single-stranded DNA in replicative intermediates, *J. Virol.* 12:1131.
- Tamanoi, F., and Stillman, B.W., 1982, Function of adenovirus terminal protein in the initiation of DNA replication, *Proc. Natl. Acad. Sci. U.S.A.* 79:2221.
- Tate, V.E., and Philipson, L., 1979, Parental adenovirus DNA accumulates in nucleosome-like structures in infected cells, *Nucleic Acids Res.* 6:2769.
- Temple, M., Antoine, G., Delius, H., Stahl, S., and Winnacker, E.L., 1981, Replication of mouse adenovirus strain FL DNA, *Virology* 109:1.
- Tibbetts, C., 1977, Physical organization of subgroup B human adenoviruses genomes, *J. Virol.* 24:564.
- Tibbetts, C., and Pettersson, U., 1974, Complementary strand-specific sequences from unique fragments of adenovirus type 2 DNA for hybridization mapping experiments, *J. Mol. Biol.* 88:767.
- Tibbetts, C., Johansson, K., and Philipson, L., 1973, Hydroxylapatite chromatography and formamide denaturation of adenovirus DNA, *J. Virol.* 12:218.
- Tibbetts, C., Pettersson, U., Johansson, K., and Philipson, L., 1974, Transcription of the adenovirus type 2 genome. I. Relationship of cytoplasmic RNA to the complementary strands of viral DNA, *J. Virol.* 13:370.
- Tokunaga, O., Shinagawa, M., and Padmanabhan, R., 1982, Physical mapping and sequence analysis at the inverted terminal repetition of adenovirus type 4, *Gene* 18:329.
- Tolun, A., Aleström, P., and Pettersson, U., 1979, Sequence of inverted terminal repetitions from different adenoviruses: Demonstration of conserved sequences and homology between SA7 and SV40 DNA, *Cell* 17:705.
- Toozé, J. (ed.), 1981, *The Molecular Biology of Tumor Viruses*, 2nd rev. ed., Part 10B, *DNA Tumor Viruses*, Cold Spring Harbor Press, Cold Spring Harbor, New York.
- Trentin, J.J., Yabe, Y., and Taylor, G., 1962, The quest for human cancer viruses, *Science* 137:835.
- van Beveren, C.P., Maat, J., Dekker, B.M.M., and van Ormondt, H., 1981, The nucleotide sequence of the gene for protein IVa2 and of the 5' leader segment of the major late mRNA's of adenovirus type 5, *Gene* 16:179.
- van den Elsen, P., de Pater, S., Houweling, A., Van der Veer, J., and van der Eb, A.J., 1982, The relationship between region E1A and E1B of human adenoviruses in cell transformation, *Gene* 18:175.
- van der Eb, A.J., and van Kesteren, L.W., 1966, Structure and molecular weight of the DNA of adenovirus type 5, *Biochim. Biophys. Acta* 129:441.
- van der Eb, A.J., van Kesteren, L.W., and van Bruggen, E.F.J., 1969, Structural properties of adenovirus DNAs, *Biochim. Biophys. Acta* 182:530.
- van der Eb, A.J., Mulder, C., Graham, F., and Houweling, A., 1977, Transformation with specific fragments of adenovirus DNA. I. Isolation of specific fragments with transforming activity of adenoviruses 2 and 5 DNA, *Gene* 2:115.
- van der Eb, A.J., van Ormondt, H., Schrier, P.I., Lupker, J.H., Jochemsen, H., van den Elsen, P.J., Deleys, R.J., Maat, J., van Beveren, C.P., Dijkema, R., and de Waard, A., 1979, Structure and function of the transforming genes of human adenoviruses and SV40, *Cold Spring Harbor Symp. Quant. Biol.* 44:383.
- van der Vliet, P.C., and Levine, A.J., 1973, DNA-binding proteins specific for cells infected by adenovirus, *Nature (London) New Biol.* 246:170.
- van der Vliet, P.C., and Sussenbach, J.S., 1975, An adenovirus type 5 gene function required for initiation of viral DNA replication, *Virology* 67:415.
- van der Vliet, P.C., Levine, A.J., Ensinger, M., and Ginsberg, H.S., 1975, Thermolabile DNA-binding proteins from cells infected with a temperature-sensitive mutant of adenovirus defective in viral DNA synthesis, *J. Virol.* 15:348.
- van der Vliet, P.C., Zandberg, J., and Jansz, H.S., 1977, Evidence for a function of the adenovirus DNA-binding protein in initiation of DNA synthesis as well as in elongation of nascent DNA chains, *Virology* 80:98.
- van Ormondt, H., Maat, J., de Waard, A., and van der Eb, A.J., 1978, The nucleotide sequence of the transforming HpaI-E fragment of adenovirus type 5 DNA, *Gene* 4:309.

- van Ormondt, H., Maat, J., and van Beveren, C.P., 1980a, The nucleotide sequence of the transforming region E-1 of adenovirus type 5 DNA, *Gene* 11:299.
- van Ormondt, H., Maat, J., and Dijkema, R., 1980b, Comparison of nucleotide sequences of the early E1A regions for subgroups A, B and C of human adenoviruses, *Gene* 12:63.
- van Wielink, P.S., 1978, A DNA-protein complex from adenovirus type 5 and its possible role in viral DNA replication, Academic thesis, State University of Utrecht, Utrecht.
- Varsanyi, T.M., Winberg, G., and Wadell, G., 1977, DNA restriction site mapping of adenovirus type 16 with BamI, EcoRI, HpaI and SalI. *FEBS Lett.* 76:151.
- Vennström, B., Pettersson, U., and Philipson, L., 1978a, Initiation of transcription in nuclei isolated from adenovirus-infected cells, *Nucleic Acids Res.* 5:205.
- Vennström, B., Pettersson, U., and Philipson, L., 1978b, Two initiation sites for adenovirus 5.5S RNA, *Nucleic Acids Res.* 5:195.
- Virtanen, A., Pettersson, U., Le Moullec, J.M., Tiollais, P., and Perricaudet, M., 1982a, Different mRNAs from the transforming region of highly oncogenic and non-oncogenic human adenoviruses, *Nature (London)* 295:705.
- Virtanen, A., Aleström, P., Persson, H., Katze, M.G., and Pettersson, U., 1982b, An adenovirus agnogene, *Nucleic Acids Res.* 10:2539.
- Vlak, J.M., Rozijn, T.H., and Sussenbach, J.S., 1975, Studies on the mechanism of replication of adenovirus DNA. IV. Discontinuous DNA chain propagation, *Virology* 63:168.
- Wadell, G., 1978, Classification of human adenoviruses by SDS-polyacrylamide gel electrophoresis of structural polypeptides, *Intervirology* 11:47.
- Weber, J., Bégin, M., and Khitto, G., 1975, Genetic analysis of adenovirus type 2. II. Preliminary phenotypic characterization of temperature-sensitive mutants, *J. Virol.* 15:1049.
- Weinmann, R., Raskas, H.J., and Roeder, R.G., 1974, Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection, *Proc. Natl. Acad. Sci. U.S.A.* 71:3436.
- Weinmann, R., Brendler, T.G., Raskas, H.J., and Roeder, R.G., 1976, Low molecular weight viral RNA's transcribed by RNA polymerase III during adenovirus 2 infection, *Cell* 7:5577.
- Westphal, H., and Lai, S.-P., 1977, Quantitative electron microscopy of adenovirus RNA, *J. Mol. Biol.* 116:525.
- Westphal, H., Meyer, J., and Maizel, J.V., Jr., 1976, Mapping of adenovirus 2 mRNA by electron microscopy, *Proc. Natl. Acad. Sci. U.S.A.* 73:2069.
- Wilson, M.C., Fraser, N.W., and Darnell, J.E., 1979, Mapping of DNA initiation sites by high doses of ultraviolet irradiation: Evidence for three independent promoters within the left 11% of the Ad-2 genome, *Virology* 94:15.
- Winberg, G., and Hammarskjöld, M.L., 1980, Isolation of DNA from agarose gels using DEAE paper: Application to restriction site mapping of adenovirus type 16 DNA, *Nucleic Acids Res.* 8:253.
- Wolfson, J., and Dressler, D., 1972, Adenovirus DNA contains an inverted terminal redundancy, *Proc. Natl. Acad. Sci. U.S.A.* 69:3054.
- Wu, M., Roberts, R.J., and Davidson, N., 1977, Structure of the inverted terminal repetition of adenovirus type 2 DNA, *J. Virol.* 21:766.
- Younghusband, H.B., and Bellett, A.J.D., 1971, Mature form of DNA from chicken embryo lethal orphan virus, *J. Virol.* 8:265.
- Zain, B.S., and Roberts, R.J., 1979, Sequences from the beginning of the fiber messenger RNA of adenovirus-2, *J. Mol. Biol.* 131:341.
- Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M., and Dunn, A.R., 1979a, Nucleotide sequence analysis of the leader segments in a cloned copy of adenovirus 2 fiber mRNA, *Cell* 16:851.
- Zain, S., Gingeras, T.R., Bullock, P., Wong, G., and Gelinas, R.E., 1979b, Determination and analysis of adenovirus-2 DNA sequences which may include signals for late mRNA processing, *J. Mol. Biol.* 135:413.
- Ziff, E.B., and Evans, R.M., 1978, Coincidence of the promoter and capped 5' terminus of RNA from the adenovirus 2 major late transcription unit, *Cell* 155:1463.
- Ziff, E.B., and Fraser, N., 1978, Adenovirus type 2 late mRNA; Structural evidence for 3'-coterminal species. *J. Virol.* 25:897.